

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ МІСЬКОГО
ГОСПОДАРСТВА імені О. М. БЕКЕТОВА

О. О. ВОРОНКОВ

КОНСПЕКТ ЛЕКЦІЙ
з курсу

ЕКОНОМЕТРИКА

*(для студентів заочної форми навчання освітньо-кваліфікаційного
рівня «бакалавр» напрямку підготовки 6.030504 - Економіка підприємства)*

Харків
ХНУМГ ім. О. М. Бекетова
2015

Воронков О. О. Конспект лекцій з курсу «Економетрика» (для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напрямку підготовки 6.030504 - Економіка підприємства) / О. О. Воронков; Харків. нац. ун-т міськ. госп-ва ім. О. М. Бекетова. - Харків: ХНУМГ ім. О. М. Бекетова, 2015. - 90 с.

Автор: канд. екон .наук О. О. Воронков

Рецензент: канд. екон .наук, доцент кафедри ЕПМГ Г. І. Базецька

*Рекомендовано кафедрою економіки підприємств міського господарства,
протокол № 1 від 27.08.2015 р)*

© О. О. Воронков, 2015

© ХНУМГ ім. О. М. Бекетова, 2015

ЗМІСТ

ВСТУП.....	4
ЗМ 1 ЕКОНОМЕТРИЧНЕ МОДЕЛЮВАННЯ. ПОБУДОВА ЗАГАЛЬНОЇ ЛІНІЙНОЇ МОДЕЛІ.....	5
Тема 1 ПРЕДМЕТ І ЗАДАЧІ ДИСЦИПЛІНИ.....	5
1.1 Предмет, методи й завдання дисципліни	5
1.2 Етапи економетричного моделювання	7
1.3 Економетрична модель. Класифікація.....	8
Тема 2 ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ.....	13
2.1 Побудова загальної лінійної моделі.....	13
2.2 Лінійна модель парної регресії.....	17
2.3 Оцінка значущості рівняння лінійної регресії	21
Тема 3 МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ.....	25
3.1 Лінійна модель множинної регресії.....	25
3.2 Оцінка значущості множинної регресії і показники якості моделі	29
3.3 Мультиколінеарність та її вплив на оцінки параметрів моделі	36
3.4 Лінійні регресійні моделі з гетероскедастичними залишками	39
3.5 Узагальнений метод найменших квадратів (УМНК).....	44
3.6 Регресійні моделі із змінною структурою. Фіктивні змінні.....	48
ЗМ 2 ЕКОНОМЕТРИЧНІ МОДЕЛІ ДИНАМІКИ. СИСТЕМА СТРУКТУРНИХ РІВНЯНЬ	53
Тема 4 ЧАСОВІ РЯДИ І ПРОГНОЗУВАННЯ	53
4.1 Загальні відомості про часові ряди і завдання їхнього аналізу	53
4.2 Автокореляція рівнів часового ряду	55
4.3 Моделювання тенденції часового ряду	57
4.4 Моделювання сезонних коливань	58
4.5 Автокореляція залишків часового ряду. Критерій Дарбіна-Уотсона.....	60
Тема 5 СИСТЕМА СТРУКТУРНИХ РІВНЯНЬ.....	62
5.1 Поняття системи структурних рівнянь.....	62
5.2 Структурна і приведена форми моделі	64
5.3 Проблема ідентифікації.....	66
5.4 Методи оцінки параметрів структурної форми моделі.....	70
ЗМ 3 МЕТОДИ АНАЛІЗУ НА ПІДСТАВІ СТАТИСТИЧНИХ РІВНЯНЬ. МОДЕЛЬ З АВТОКОРЕЛЬОВАНИМИ ЗАЛИШКАМИ. МОДЕЛІ РОЗПОДІЛЕНОГО ЛАГУ	73
Тема 6 НЕЛІНІЙНІ ОДНОФАКТОРНІ МОДЕЛІ	73
6.1 Види нелінійних моделей парної регресії.....	73
6.2 Моделі нелінійні щодо пояснюючих змінних	74
6.3 Моделі нелінійні за оцінюваними параметрами.....	75
6.4 Виробнича функція Кобба-Дугласа	78
Тема 7 РЕГРЕСІЙНІ ДИНАМІЧНІ МОДЕЛІ.....	80
7.1 Причини виникнення лагових ефектів в економетричних моделях....	80
7.2 Методи оцінки параметрів з урахуванням лагових ефектів.....	82
СПИСОК ДЖЕРЕЛ	89

ВСТУП

Курс «Економетрика» є нормативною дисципліною в навчальному плані за напрямком «Економіка і підприємництво» для кваліфікаційного рівня «Бакалавр». Обсяг курсу становить 108 академічних годин або 3,0 кредити. При вивченні за заочною формою обсяг аудиторних занять становить 12 годин (6 годин лекцій і 6 годин практичних занять), на самостійну роботу студента припадає 96 годин. Програма курсу містить 3 змістових модулі: «Економетричне моделювання. Побудова загальної лінійної моделі», «Економетричні моделі динаміки. Системи структурних рівнянь» і «Методи аналізу на підставі статистичних рівнянь. Модель з автокорельованими залишками. Моделі розподіленого лага», відповідно до яких виконують проміжний контроль знань. Підсумковий контроль знань (залік) проводять в усній формі. У процесі вивчення курсу студенти повинні виконати контрольну роботу.

Метою вивчення дисципліни «Економетрика» є освоєння методів побудови й оцінки параметрів залежностей, що характеризують кількісні взаємозв'язки в економічних процесах з метою їхнього аналізу й прогнозування.

У результаті вивчення курсу студенти повинні вміти вибрати клас економетричної моделі для досліджуваного процесу, оцінити параметри й знайти їхні інтервальні оцінки, перевірити значущість моделі, а також на основі прогнозних значень змінних, оцінивши їхню вірогідність, приймати рішення в завданнях управління економічними процесами на різних ієрархічних рівнях.

При викладенні навчального матеріалу передбачають, що студент володіє основами теорії імовірностей і математичної статистики, а також лінійної алгебри в обсязі курсу математики для економістів.

Основна увага в курсі приділяється розгляду лінійних економетричних моделей як найбільш простих і таких, що володіють найменшим ризиком одержання значних помилок прогнозу. З тієї ж причини вивчення часових рядів обмежується розглядом в основному стаціонарних рядів.

Економіка все більше стає однією з найбільш математизованих наук. Досягнення сучасної економічної науки висувають нові вимоги до вищої професійної освіти економістів. У період становлення ринкової економіки зростає роль економетричних методів. Без знання цих методів неможливі ні дослідження й теоретичне узагальнення емпіричних залежностей економічних змінних, ні побудова скільки-небудь надійного прогнозу в банківській справі, фінансах або бізнесі.

ЗМ 1 ЕКОНОМЕТРИЧНЕ МОДЕЛЮВАННЯ. ПОБУДОВА ЗАГАЛЬНОЇ ЛІНІЙНОЇ МОДЕЛІ

Тема 1 ПРЕДМЕТ І ЗАДАЧІ ДИСЦИПЛІНИ

1.1 Предмет, методи і завдання дисципліни

Економетрика - одна з базових дисциплін економічної освіти в усьому світі. Існують різні варіанти визначення Економетрики, в числі яких - розширені, при яких до Економетрики відносять все, що пов'язано з вимірюваннями в економіці, і вузько інструментально орієнтовані, при яких під економетрикою розуміють певний набір математико-статистичних засобів, що дозволяють верифікувати модельні співвідношення між аналізованими економічними показниками.

Найбільш точно пояснив сутність Економетрики один із засновників цієї науки норвезький вчений Р.Фріш, що і ввів цю назву в 1926 р.: «Економетрика - це не є те саме, що економічна статистика. Вона не ідентична й тому, що ми називаємо економічною теорією, хоча значна частина цієї теорії має кількісний характер. Економетрика не є синонімом застосування математики до економіки. Як показує досвід, кожна з трьох відправних точок - статистика, економічна теорія та математика - необхідна, але не достатня умова для розуміння кількісних співвідношень в сучасному економічному житті. Це єдність усіх трьох складових. І ця єдність утворює економетрику».

Економетрика - це самостійна наукова дисципліна, що поєднує сукупність теоретичних результатів, прийомів, методів і моделей, призначених для того, щоб на базі економічної теорії, економічної статистики і апарату математичної статистики надавати конкретне кількісне вираження загальним (якісним) закономірностям, що обумовлені економічною теорією.

Економетричні методи склалися з урахуванням особливостей економічних явищ і процесів, які спотворюють результати застосування класичних статистичних методів. Такими особливостями, зокрема, є асиметричність і мультиколінеарність зв'язків, ефект гетероскедастичності, автокореляція, помилкова кореляція, наявність лагів. Сутність використаних термінів буде розкрита у процесі вивчення цього курсу.

Таким чином, економетрика поєднує сукупність методів і моделей, що дозволяють на базі економічної теорії, економічної статистики і математико-статистичного апарату надавати кількісні вираження якісним залежностям. Основні результати економічної теорії мають якісний характер, а Економетрика вносить до них емпіричний зміст. Математична економіка виражає економічні закони у вигляді математичних співвідношень, а економетрика здійснює досліду перевірку цих законів. Економічна статистика дає інформаційне забезпечення досліджуваного процесу у вигляді вихідних (оброблених) статистичних даних і економічних показників, а економетрика, використовуючи традиційні

математико-статистичні та спеціально розроблені методи, аналізує кількісні взаємозв'язки між цими показниками.

Багато базових понять економетрики мають два визначення - економічне та математичне. Подібна подвійність має місце і у формулюваннях результатів. Характер наукових праць з економетрики варіюється від «класичних» економічних робіт, в яких майже не використовується математичний апарат, до солідних математичних праць, що використовують досить тонкий апарат сучасної математики.

Економічна складова економетрики, безумовно, є первинною. Саме економіка визначає постановку задачі та вихідні передумови, а результат, сформований математичною мовою, складає інтерес лише в тому випадку, якщо вдається його економічна інтерпретація. У той самий час велика кількість економетричних результатів має характер математичних тверджень (теорем).

Широкому впровадженню економетричних методів сприяла поява в другій половині XX ст. електронних обчислювальних машин, зокрема персональних комп'ютерів. Комп'ютерні економетричні пакети зробили ці методи доступнішими та наочнішими, тому що найбільш трудомістку (рутинну) роботу з розрахунку різних статистик, параметрів, характеристик, побудови таблиць і графіків в основному став виконувати комп'ютер, а економетристу залишилася головним чином творча робота: постановка задачі, вибір відповідної моделі та методу її розв'язання, інтерпретація результатів.

Економетричне моделювання реальних соціально-економічних процесів і систем зазвичай переслідує одну з двох кінцевих прикладних цілей: прогноз економічних і соціально-економічних показників, що характеризують стан і розвиток аналізованої системи, імітацію різних можливих сценаріїв соціально-економічного розвитку аналізованої системи (різноманітні сценарні розрахунки, ситуаційне моделювання).

При постановці задач економетричного моделювання треба визначити їхній ієрархічний рівень і профіль. Аналізовані задачі можуть належати до макро- (країна, міждержавний аналіз), мезо- (регіони усередині країни) і мікро- (підприємства, фірми, родини) рівнів і бути спрямованими на вирішення питань різного профілю інвестиційної, фінансової або соціальної політики, ціноутворення, розподільних стосунків та ін.

Загальним моментом для будь-якої економетричної моделі є розбивка залежної змінної на дві частини - пояснену та випадкову. Сформулюємо задачу моделювання в самому загальному вигляді: на підставі експериментальних даних визначити пояснену частину і, розглядаючи випадкову складову як випадкову величину, отримати (можливо, після певних припущень) оцінки параметрів її розподілу.

Таким чином, економетрична модель має такий вигляд:

Спостережене значення залежної змінної	=	Пояснена частина, залежна від значень пояснюючих змінних	+	Випадкова складова
$Y = f(x) + \varepsilon.$				
(1.1)				

Нехай отримано наступний вираз для поясненої частини змінної Y :

$$y = b + b_1x_1 + b_2x_2.$$

Очевидно, що він дає уявлення про те, як саме формується розглянута економічна змінна y і дає можливість виявити вплив на неї кожної з пояснюючих змінних x_1 і x_2 . В цьому випадку пояснювальна змінна y дорівнює b при $x_1=0$ і $x_2=0$. За рахунок збільшення x_1 на 1 одиницю вона збільшується на b_1 , а за рахунок збільшення x_2 на 1 одиницю зростає на b_2 . Найбільше важливо те, що отриманий вираз дозволяє прогнозувати значення поясненої частини змінної Y , якщо відомі його параметри b , b_1 і b_2 .

1.2 Етапи економетричного моделювання

Прийнято виділяти шість основних етапів економетричного моделювання: постановочний, апріорний, етап параметризації, інформаційний, етапи ідентифікації і верифікації моделі.

Зупинимось докладніше на кожному з цих етапів і розглянемо проблеми, пов'язані з їхньою реалізацією.

1-й етап (постановочний). На даному етапі формують мету дослідження і визначають набір економічних змінних моделі. При виборі економічних змінних необхідно здійснювати теоретичне обґрунтування кожної змінної. При цьому рекомендують, щоб число їх було не дуже великим і, як мінімум, у кілька разів менше числа спостережень. Пояснюючі змінні не повинні бути пов'язані між собою функціональною або кореляційною залежністю, тому що це може призвести до неможливості оцінки параметрів моделі або до одержання нестійких оцінок, що не мають реального смислу, тобто до явища мультиколінеарності. Визначальним при включенні в модель тих або інших змінних є економічний (якісний) аналіз досліджуваного об'єкта.

2-й етап (апріорний). На даному етапі проводять аналіз сутності досліджуваного об'єкта, формування й формалізація апріорної інформації.

3-й етап (параметризація). На даному етапі здійснюють безпосередньо моделювання, тобто вибір загального виду моделі, виявлення у зв'язків, які до неї входять. Основна задача, розв'язувана на цьому етапі, - вибір виду функції $f(x)$. Досить важливою проблемою на цьому етапі економетричного моделювання є проблема специфікації моделі, зокрема, вираження в математичній формі виявлених зв'язків і співвідношень; установлення складу екзогенних (незалежних) і ендогенних (залежних) змінних, формулювання вихідних передумов і обмежень моделі. Від того, наскільки вдало вирішена проблема специфікації моделі, у значній мірі залежить успіх всього економетричного моделювання.

4-й етап (інформаційний). На даному етапі здійснюють збір необхідної статистичної інформації – спостережуваних значень економічних змінних

$$(x_{i1}, x_{i2}, \dots, x_{ip}; y_{i1}, y_{i2}, \dots, y_{iq}), i=1, n.$$

Тут можуть бути спостереження, отримані як за участю дослідника, так і без його участі (в умовах активного або пасивного експерименту).

5-й етап (ідентифікація моделі). Тут здійснюють статистичний аналіз моделі й оцінка її параметрів.

6-й етап (верифікація моделі). На даному етапі проводять перевірку істинності, тобто адекватності моделі. З'ясовують, наскільки вдало вирішені проблеми специфікації, ідентифікації та ідентифікованості моделі, яка точність розрахунків, зроблених на підставі даної моделі, і, в остаточному підсумку, наскільки відповідає побудована модель реальному економічному об'єкту або процесу, який моделюють. Варто помітити, що коли є статистичні дані, які характеризують економічний об'єкт, що моделюють, у поточний і попередній моменти часу, то для верифікації моделі, побудованої для прогнозу, досить зрівняти реальні значення змінних в наступні моменти часу з відповідними їм значеннями, отриманими на основі розглянутої моделі за даними попередніх моментів.

Розглянутий поділ економетричного моделювання на окремі етапи носить певною мірою умовний характер, тому що ці етапи можуть перетинатися, взаємно доповнювати один одного та ін.

1.3 Економетрична модель. Класифікація

Нехай є p пояснюючих змінних (факторів) x_1, x_2, \dots, x_p і залежна змінна Y . Змінна Y є випадковою величиною, що має при заданих значеннях факторів x_j деякий розподіл. Звичайно припускають, що умовні розподіли Y при кожному припустимому значенні факторів x_j - нормальні.

Пояснюючі змінні x_j ($j = 1, \dots, p$) можуть вважатися як випадковими, так і детермінованими, тобто приймаючими цілком певні значення.

Класична економетрична модель розглядає пояснюючі змінні x_j як детерміновані.

Пояснена частина Y_e в кожному разі являє собою функцію від значень факторів - пояснюючих змінних:

$$Y_e = f(x_1, x_2, \dots, x_p).$$

Таким чином, економетрична модель має вигляд

$$Y = f(x_1, x_2, \dots, x_p) + \varepsilon.$$

Найбільш природним вибором поясненої частини випадкової величини Y є її середнє значення - умовне математичне сподівання $M_x(Y)$, отримане при даному наборі значень пояснюючих змінних (x_1, x_2, \dots, x_p) .

Рівняння $M_x(Y) = f(x_1, x_2, \dots, x_p)$ називається рівнянням регресії.

При такому природному виборі поясненої частини економетрична модель має вигляд

$$Y = M_x(Y) + \varepsilon, \quad (1.2)$$

де ε - випадкова величина, називана збуренням або помилкою. В курсі математичної статистики рівняння (1.2) називають рівнянням регресійної моделі.

Відзначимо, що економетрична модель не обов'язково є регресійною, тобто пояснена частина не завжди являє собою умовне математичне сподівання

залежної змінної. Однією з можливих причин того, що економетрична модель не є регресійною, є наявність систематичних помилок виміру пояснюючих змінних. З математичної точки зору регресійні моделі виявляються істотно більш простим об'єктом, ніж економетрична модель загального типу (1.1).

Щоб одержати досить достовірні й інформативні дані щодо розподілу будь-якої випадкової величини, необхідно мати вибірку її спостережень досить великого обсягу. Така вибірка спостережень залежної змінної Y і пояснюючих змінних x_j ($j = 1, \dots, p$) є відправною точкою будь-якого економетричного дослідження. Такі вибірки являють собою набори значень $(x_{i1}, \dots, x_{ip}; y_i)$, де $i = 1, n$; p - кількість пояснюючих змінних, n - число спостережень.

Як правило, число спостережень n досить велике й значно перевищує число p пояснюючих змінних. Проте проблема полягає в тому, що спостереження y_i , розглянуті в різних вибірках як випадкові величини Y_i і одержувані при різних наборах значень пояснюючих змінних x_j , мають в загальному випадку різний розподіл. Це означає, що для кожної випадкової величини Y_i ми маємо всього лише одно спостереження. Зрозуміло, на підставі одного спостереження не можна зробити адекватний висновок щодо розподілу випадкової величини, і потрібні додаткові припущення.

Прийнято виділяти три основних класи моделей, які застосовують для аналізу або прогнозу, це:

- регресійні моделі з одним рівнянням,
- моделі часових рядів,
- системи одночасних рівнянь.

Вибір моделі залежить від виду вихідних даних, на підставі яких будують економетричну модель, а також від характеру досліджуваного економічного процесу.

В класичному курсі Економетрики розглядають два типи вибірових даних: просторові дані й часові ряди.

Просторова вибірка. Прикладом просторових даних є, наприклад, набір відомостей (обсяг виробництва, кількість працівників, дохід та ін.) з різних фірм у той самий момент часу (просторовий зріз). Іншим прикладом можуть бути дані з курсів покупки/продажу наявної валюти в якийсь день в обмінних пунктах міста та ін.

В економіці під просторовою вибіркою розуміють набір показників економічних змінних, отриманий у цей момент часу. Для економетриста, однак, таке визначення не дуже зручне (через неоднозначність поняття «момент часу». Це може бути і день, і тиждень, і рік). Очевидно, про просторову вибірку має сенс говорити в тому випадку, якщо всі спостереження отримані приблизно в незмінних умовах, тобто являють собою набір незалежних вибірових даних з деякої генеральної сукупності.

Будемо називати просторовою вибіркою серію з n незалежних спостережень p -мірної випадкової величини $(x_{i1}, \dots, x_{ip}; y_i)$. При цьому надалі можна не розглядати X_j як випадкові величини. В цьому випадку різні випадкові величи-

ни Y_i виявляються між собою незалежними, що спричиняє некорельованість їхніх збурень, тобто коефіцієнт кореляції між збуреннями ε_i і ε_j

$$r(\varepsilon_i, \varepsilon_j) = 0 \quad \text{при } i=j. \quad (1.3)$$

Умова (1.3) істотно спрощує модель та її статистичний аналіз.

Як визначити, чи є вибірка серією незалежних спостережень? На це питання немає однозначної відповіді. Формальне визначення незалежності випадкових величин, як правило, виявляється реально неперевіреною. Зазвичай як незалежні приймають величини, що зв'язані причинно. Однак на практиці далеко не завжди питання про незалежність виявляється безперечним.

Економетрична модель, побудована на основі просторової вибірки експериментальних даних (x_i, y_i) є регресійною моделлю з одним рівнянням

$$y_i = f(x_i) + \varepsilon_i, \quad i=1 \dots n, \quad (1.4)$$

де помилки регресії задовольняють умовам

$$M(\varepsilon_i) = 0, \quad (1.5)$$

$$r(\varepsilon_i, \varepsilon_j) = 0, \quad (1.6)$$

$$D(\varepsilon_i) = \sigma_i^2. \quad (1.7)$$

Що стосується умови (1.7), то тут можливі два випадки:

а) $\sigma_i^2 = \sigma_j^2$ при всіх i та j . Властивість сталості дисперсій помилок регресії називається **гомоскедастичністю**. В цьому випадку розподіли випадкових величин Y_i відрізняються тільки значенням математичного сподівання (поясненої частини);

б) $\sigma_i^2 \neq \sigma_j^2$. У цьому випадку має місце **гетероскедастичність** моделі. Гетероскедастичність «псує» багато результатів статистичного аналізу і, як правило, вимагає усунення.

В деяких випадках гетероскедастичність моделі очевидна, однак у більшості випадків потрібне застосування методів математичної статистики для прийняття рішення про те, який тип моделі варто розглядати.

Залежно від вигляду функції $f(x_i)$ моделі ділять на лінійні і нелінійні. Наприклад, можна досліджувати попит на морозиво як функцію від часу, від температури повітря, від середнього рівня доходів або залежність зарплати від віку, статі, рівня освіти, стажу роботи і т.п. Область застосування таких моделей, заснованих на просторових даних, навіть лінійних, значно ширша, ніж моделей часових рядів.

Часовий (динамічний) ряд. Прикладами часових даних можуть бути щоквартальні дані з інфляції, середньої заробітної плати, національного доходу, грошової емісії за останні роки або, наприклад, щоденний курс долара США на МВБ, ціни ф'ючерсних контрактів на поставку долара США та ін.

Відмінною рисою часових даних є те, що вони природно впорядковані за часом, крім того, спостереження в близькі моменти часу часто бувають залежними.

Часовим (динамічним) рядом називають вибірку спостережень, в якій важливі не тільки самі спостережувані значення випадкових величин, але й порядок їхнього проходження один за одним. Найчастіше впорядкованість зумовле-

на тим, що експериментальні дані являють собою серію спостережень однієї і тієї ж випадкової величини в послідовні моменти часу. В цьому випадку динамічний ряд називають часовим рядом. При цьому передбачається, що тип розподілу спостережуваної випадкової величини залишається тим самим (наприклад, нормальним), але параметри його міняються залежно від часу.

Моделі часових рядів, як правило, є складніше моделей просторової вибірки, тому що спостереження у випадку часового ряду, загалом кажучи, не є незалежними, а це значить, що помилки регресії можуть корелювати одна з одною, тобто умова (1.3) не виконується. Це значно ускладнює статистичний аналіз моделі.

Відзначимо, що маючи тільки ряд спостережень без розуміння їхньої природи, неможливо визначити, маємо ми справу з просторовою вибіркою або часовим рядом.

До класу моделей часових рядів відносять моделі тренду, сезонності, а також тренду й сезонності.

Модель тренду має вигляд:

$$Y(t) = T(t) + \varepsilon_t, \quad (1.8)$$

де $T(t)$ - часовий тренд заданого параметричного виду (наприклад, лінійний) $T(t) = a + bt$;

ε - випадкова (стохастична) компонента.

Модель сезонності має вигляд:

$$Y(t) = S(t) + \varepsilon_t, \quad (1.9)$$

де $S(t)$ - періодична (сезонна) компонента.

Модель тренду й сезонності може бути адитивною або мультиплікативною моделлю:

$$y(t) = T(t) + S(t) + \varepsilon_t \quad - \text{аддитивна};$$

$$y(t) = T(t) * S(t) + \varepsilon_t \quad - \text{мультиплікативна},$$

де $T(t)$ - часовий тренд заданого параметричного виду;

$S(t)$ - періодична (сезонна) компонента.

До моделей часових рядів відносять множину більш складних моделей, таких як моделі адаптивного прогнозу, моделі авторегресії і ковзного середнього та ін. Їхньою загальною рисою є те, що вони пояснюють поведінку часового ряду, виходячи тільки з його попередніх значень. Такі моделі можуть застосовуватися, наприклад, для вивчення й прогнозування обсягу продажів авіаквитків, попиту на морозиво, короткострокового прогнозу процентних ставок та ін.

Моделі, які описуються системами рівнянь, називаються **системами одночасних рівнянь**. Системи можуть складатися з тотожностей і регресійних рівнянь, кожне з яких може, крім пояснюючих змінних, містити в собі пояснювальні змінні з інших рівнянь системи. Таким чином, тут ми маємо набір пояснювальних змінних, зв'язаних через рівняння системи. Класичним прикладом системи одночасних рівнянь є модель попиту Q^d і пропозиції Q^s . Коли попит на товар визначають його ціною P і доходом споживача I , пропозиція товару – його ціною P і досягається рівновага між попитом та пропозицією:

$$Q^s = \alpha_1 + \alpha_2 P + \varepsilon_1 \quad (\text{пропозиція}), \quad (1.10)$$

$$Q^d = \beta_1 + \beta_2 P + \beta_3 I + \varepsilon_2 \quad (\text{попит}), \quad (1.11)$$

$$Q^d = Q^s \quad (\text{рівновага}). \quad (1.12)$$

В цій системі екзогенною (незалежною) змінною є дохід споживача I , а ендогенними (залежними) – попит (пропозиція) товару $Q^d = Q^s = Q$ і ціна товару (ціна рівноваги) P .

В іншій моделі попиту та пропозиції як пояснююча пропозицію змінна може бути не тільки ціна товару P в даний момент часу t , але й ціна товару в попередній момент часу $t-1$, тобто лагова ендогенна змінна:

$$Q_t^s = \alpha_1 + \alpha_2 P_t + \alpha_3 P_{t-1} + \varepsilon_1, \quad (1.13)$$

Системи одночасних рівнянь вимагають застосування більш складного математичного апарату. Вони можуть використовуватися для моделей стронової економіки та ін.

Таким чином, економетрична модель дозволяє пояснити поведінку ендогенних змінних залежно від значень екзогенних і лагових ендогенних змінних, інакше кажучи, залежно від визначених змінних.

Завершуючи розгляд поняття економетричної моделі, слід зазначити наступне. Не усяка економіко-математична модель, що представляє математико-статистичний опис досліджуваного економічного об'єкта, може вважатися економетричною. Вона стає економетричною тільки в тому випадку, якщо буде відбивати цей об'єкт на основі емпіричних (статистичних) даних, що характеризують саме його.

Запитання для самоперевірки

1. Наведіть визначення економетрики. За якими ознаками економетрику визначають як відокремлену науку? З якими науками пов'язана економетрика?
2. З якою метою проводять економетричні дослідження?
3. Охарактеризуйте етапи економетричного дослідження. Які питання вирішує економетрика? Яку роль відіграє математична статистика у формуванні економетричних методів?
4. Які типи даних використовують у економетричних дослідженнях. Які виникають проблеми з даними?
5. Поясніть складові загальної економетричної моделі. Поясніть відмінності між ендогенними та екзогенними змінними.
6. У чому полягає ідентифікація та верифікація економетричної моделі?
7. Поясніть терміни «регресія», «умовне математичне сподівання» та «збурення». Які класи моделей використовують для аналізу або прогнозу в економетриці?
8. Які властивості повинна мати регресійна модель з одним рівнянням, побудована на основі просторової вибірки?
9. Поясніть терміни «гомоскедастичність» та «гетероскедастичність».
10. Яку вибірку спостережень називають часовим рядом?
11. Які економетричні моделі належать до систем одночасних рівнянь?

Тема 2 ПАРНИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

2.1 Побудова загальної лінійної моделі

В практиці економічних досліджень наявні дані не завжди можна вважати вибіркою з багатомірної нормальної сукупності, коли одна з розглянутих змінних не є випадковою або коли лінія регресії явно не пряма та ін. В цих випадках намагаються визначити криву (або поверхню), що дає найкраще (у сенсі методу найменших квадратів) наближення до вихідних даних. Відповідні методи наближення одержали назву регресійного аналізу.

Методи й моделі регресійного аналізу займають центральне місце в математичному апараті Економетрики.

Задачами регресійного аналізу є встановлення форми залежності між змінними, оцінка функції регресії, оцінка невідомих значень (прогноз значень) залежної змінної.

В економіці в більшості випадків між змінними величинами існують залежності, в яких кожному значенню однієї змінної відповідає не якийсь певний, а множина можливих значень іншої змінної. Інакше кажучи, кожному значенню однієї змінної відповідає певний (умовний) розподіл іншої змінної. Така залежність одержала назву статистичної (або стохастичної, імовірнісної).

Виникнення поняття статистичного зв'язку зумовлено тим, що залежна змінна піддана впливу ряду неконтрольованих або неврахованих факторів, а окрім того тим, що вимір значень змінних неминує супроводжуватися деякими випадковими помилками. Прикладом статистичного зв'язку є залежність урожайності від кількості внесених добрив, продуктивності праці на підприємстві від його енергооснащеності та ін.

В силу неоднозначності статистичної залежності між Y і X для дослідника, зазвичай, становить інтерес усереднена за X схема залежності, тобто закономірність у зміні умовного математичного сподівання $M_x(Y)$ (математичного сподівання випадкової змінної Y , обчисленого в припущенні, що змінна X прийняла значення x) залежно від x .

Якщо залежність між двома змінними така, що кожному значенню однієї змінної відповідає певне умовне математичне сподівання (середнє значення) іншої, то така статистична залежність називається *кореляційною*. Інакше кажучи, кореляційною залежністю між двома змінними називають функціональну залежність між значеннями однієї з них і умовним математичним сподіванням іншої.

Кореляційна залежність може бути представлена у вигляді

$$M_x(Y) = \varphi(x). \quad (2.1)$$

В регресійному аналізі розглядається однобічна залежність випадкової змінної Y від однієї (або кількох) невідповідної незалежної змінної X . Така залежність може виникнути, наприклад, у випадку, коли при кожному фіксованому значенні X відповідні значення Y піддані випадковому розкиду за рахунок

дії ряду неконтрольованих факторів. Таку залежність Y від X називають **регресійною**. При цьому залежну змінну Y називають також функцією відгуку, пояснювальною, вихідною, результуючою, ендогенною змінною, результативною ознакою, а незалежну змінну X - пояснюючою, вхідною, предикторною, екзогенною змінною, фактором, регресором, факторною ознакою.

Рівняння (2.1) називається **рівнянням регресії**, функція $\varphi(x)$ - функцією регресії, а її графік - **лінією регресії**.

Для точного опису рівняння регресії необхідно знати умовний закон розподілу залежної змінної Y за умови, що змінна X прийме значення x ($X=x$). У статистичній практиці таку інформацію одержати, як правило, не вдається, тому що зазвичай дослідник має лише вибірку пар значень (x_i, y_i) обмеженого обсягу n . У цьому випадку мова може йти про оцінку (наближений вираз, апроксимацію) за вибіркою функції регресії. Такою оцінкою є вибіркова лінія регресії:

$$\hat{y} = \hat{\varphi}(x, b_0, b_1, \dots, b_p), \quad (2.2)$$

де y - умовна (групова) середня змінної Y при фіксованому значенні змінної $X=x$;

b_0, b_1, \dots, b_p - параметри кривої.

Рівняння (2.2) називають вибірковим рівнянням регресії.

При правильно визначеній апроксимуючій функції $\varphi(x, b_0, b_1, \dots, b_p)$ із збільшенням обсягу вибірки ($n \rightarrow \infty$) вона буде збігатися за імовірністю до функції регресії $\varphi(x)$.

Модель парної регресії називають простою (з двома змінними) економетричною моделлю. Її застосовують, якщо завдання полягає у визначенні або в оцінці залежності пояснювальної змінної Y від одного фактору (пояснюючої змінної) X . При цьому вихідні дані являють собою два набори значень (вектори) $x = (x_1, \dots, x_n)$ і $y = (y_1, \dots, y_n)$. Метою є одержання аналітичної залежності пояснювальної змінної Y від пояснюючої змінної x , що щонайкраще відповідає вихідним даним.

Таким чином, парна регресія являє собою регресію між двома змінними – y та x , тобто модель виду:

$$y = \hat{f}(x), \quad (2.3)$$

де y - залежна змінна (результативна ознака);

x - незалежна, або пояснююча, змінна (ознака-фактор).

Знак « \wedge » означає, що між змінними x та y немає строгої функціональної залежності, тому практично в кожному окремому випадку величина y складається з двох доданків:

$$y = \hat{y}_x + \varepsilon, \quad (2.4)$$

де y – фактичне значення результативної ознаки;

\hat{y}_x – теоретичне значення результативної ознаки, знайдене виходячи з рівняння регресії;

ε – випадкова величина, що характеризує відхилення реального значення результативної ознаки від теоретичного, знайденого за рівнянням регресії.

Випадкову величину ε називають також **збуренням**. Вона включає вплив неврахованих у моделі факторів, випадкових помилок і особливостей виміру. Її присутність в моделі породжено трьома джерелами: специфікацією моделі, вибіркоvim характером вихідних даних і особливостями вимірювання змінних.

В регресійному аналізі розглядають однобічну залежність випадкової змінної Y від однієї (або кількох) не випадкової незалежної змінної X .

Від правильно обраної специфікації моделі залежить величина випадкових помилок: вони тим менші, чим принаймні теоретичні значення результативної ознаки \hat{y}_x відрізняються від фактичних даних y .

До помилок специфікації відносять неправильний вибір тієї або іншої математичної функції для \hat{y}_x і не врахування в рівнянні регресії будь-якого істотного фактору, тобто використання парної регресії замість множинної.

Поряд з помилками специфікації можуть мати місце помилки вибірки, які звичайно обумовлені неоднорідністю даних у вихідній статистичній сукупності. Це, як правило, буває при вивченні економічних процесів. Якщо сукупність неоднорідна, то рівняння регресії не має практичного змісту. Для одержання доброго результату звичайно виключають із сукупності одиниці з аномальними значеннями досліджуваних ознак.

Використання часової інформації також являє собою вибірку з усієї множини хронологічних дат. Змінивши часовий інтервал, можна одержати інші результати.

Найбільшу небезпеку в практичному використанні методів регресії представляють помилки виміру. Якщо помилки специфікації можна зменшити, змінюючи форму моделі (вид математичного виразу), а помилки вибірки - збільшуючи обсяг вихідних даних, то помилки виміру практично зводять нанівець усі зусилля з кількісної оцінки зв'язку між ознаками.

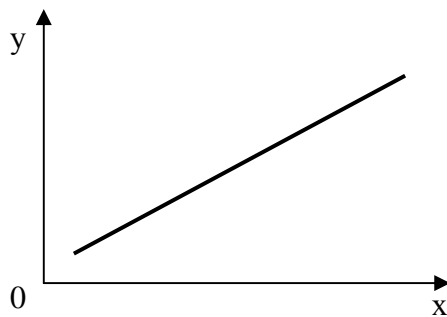
Особливо велику роль відіграють помилки виміру при дослідженні на макрорівні. Так, у дослідженнях попиту і споживання як пояснюючу змінну широко використовують «дохід населення». Разом з тим, статистичний вимір величини доходу пов'язаний з низкою труднощів і не позбавлений можливих помилок, наприклад, в результаті наявності схованих доходів.

Припускаючи, що помилки виміру зведені до мінімуму, основну увагу в економетричних дослідженнях приділяють помилкам специфікації моделі.

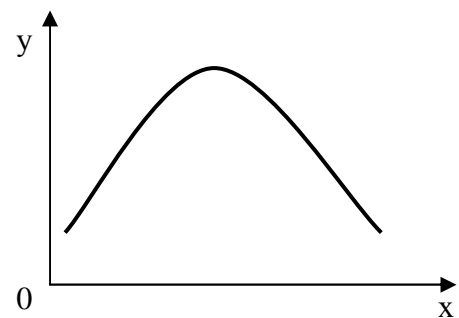
В парній регресії вибір вигляду математичної функції $\hat{y}_x = f(x)$ можна здійснити трьома методами:

- графічним;
- аналітичним, тобто виходячи з теорії досліджуваного взаємозв'язку;
- експериментальним.

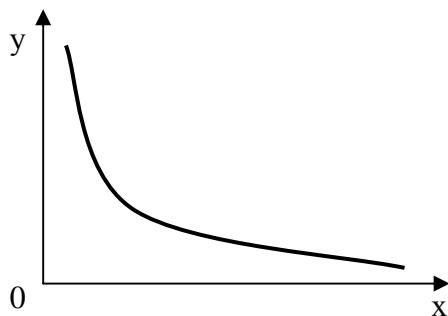
При вивченні залежності між двома ознаками графічний метод підбору виду рівняння регресії досить наочний. Він заснований на побудові поля кореляції. Основні типи кривих, які використовують при кількісній оцінці зв'язків, зображені на рисунку 2.1.



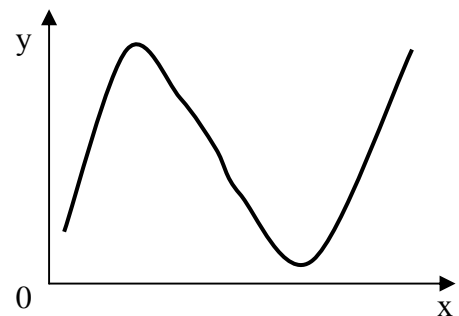
$$\hat{y}_x = a + bx$$



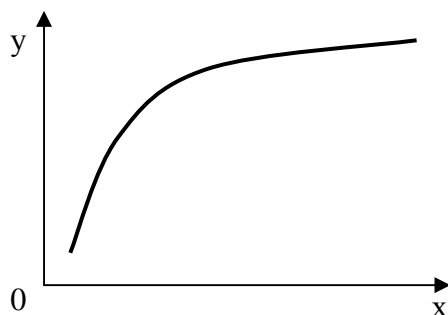
$$\hat{y}_x = a + bx + cx^2$$



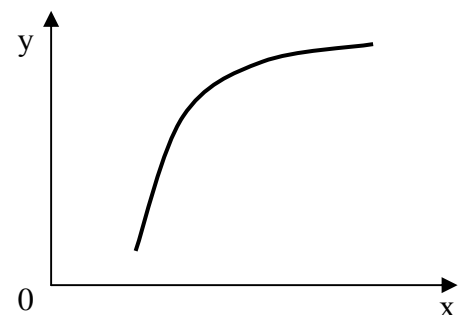
$$\hat{y}_x = a + \frac{b}{x}$$



$$\hat{y}_x = a + bx + cx^2 + dx^3$$



$$\hat{y}_x = a + x^b$$



$$\hat{y}_x = a + b^x$$

Рисунок 2.1 - Основні типи кривих, які використовують при кількісній оцінці зв'язків між двома змінними

Значну зацікавленість виявляє аналітичний метод вибору типу рівняння регресії. Він заснований на вивченні матеріальної природи зв'язку досліджуваних ознак.

При обробці інформації на комп'ютері вибір виду рівняння регресії зазвичай здійснюють експериментальним методом, тобто шляхом порівняння величини залишкової дисперсії $\sigma_{зал}^2$, яку розраховують для різних моделей.

Якщо рівняння регресії охоплює всі точки кореляційного поля, що можливо тільки при функціональному зв'язку, коли всі точки лежать на лінії регресії $\hat{y}_x = f(x)$, то фактичні значення результативної ознаки збігаються з теоретичними $y = \hat{y}_x$, тобто вони цілком зумовлені впливом фактору x . В цьому випадку залишкова дисперсія $\sigma_{зал}^2 = 0$.

В практичних дослідженнях, як правило, має місце певне розсіювання точок кореляційного поля щодо лінії регресії. Воно зумовлене впливом інших факторів, що не враховуються в рівнянні регресії. Іншими словами, мають місце відхилення фактичних даних від теоретичних $(y - \hat{y}_x)$. Величина цих відхилень лежить в основі розрахунку залишкової дисперсії:

$$\sigma_{зал}^2 = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_x)^2. \quad (2.5)$$

Чим меншою є величина залишкової дисперсії, тим менший вплив факторів, що не враховуються в рівнянні регресії, і тим краще рівняння регресії підходить до вихідних даних.

Вважають, що число спостережень має в 7-8 разів перевищувати число параметрів при змінній x , що розраховуються. Це означає, що шукати лінійну регресію, маючи менше за 7 спостережень, взагалі не має рації. Якщо вигляд функції ускладнюється, то необхідне збільшення обсягу спостережень, тому що кожний параметр при x повинен розраховуватися хоча б за 7 спостереженнями. Виходить, якщо ми обираємо криву другого ступеня (параболу) $\hat{y}_x = a + bx + cx^2$, то потрібен обсяг інформації понад 14 спостережень.

2.2 Лінійна модель парної регресії

Найбільш добре вивчені лінійні регресійні моделі, що задовольняють умовам (1.5), (1.6) і властивості сталості дисперсії помилок регресії, - їх називають **класичними моделями**.

Помітимо, що умовам класичної регресійної моделі задовольняють і гомоскедастична модель просторової вибірки, і модель часового ряду, спостереження якого не корелюють, а дисперсії постійні. З математичної точки зору вони дійсно нерозрізніювані (хоча можуть значно розрізнятися економічні інтерпретації отриманих математичних результатів). Нехай визначено характер експериментальних даних і виділено певний набір пояснюючих змінних. Для того щоб знайти пояснену частину, тобто величину $M_x(Y)$, потрібне знання умовних розподілів випадкової величини Y . На практиці це майже ніколи не має місця, тому точне знаходження поясненої частини неможливе.

В таких випадках застосовують стандартну процедуру згладжування експериментальних даних. Ця процедура складається з двох етапів:

- визначають параметричне сімейство, до якого належить шукана функція $M_x(Y)$ (розглянута як функція від значень пояснюючих змінних X). Це може бути множина лінійних функцій, степеневих функцій та ін. (рис. 2.1);

- визначають оцінки параметрів цієї функції за допомогою одного з методів математичної статистики.

Формально ніяких способів вибору параметричного сімейства не існує. Проте у переважній більшості випадків економетричні моделі вибирають лінійними.

Окрім цілком очевидної переваги лінійної моделі - її відносної простоти, - для такого вибору є, принаймні, дві істотні причини. Перша причина: якщо випадкова величина (X, Y) має спільний нормальний розподіл, тоді, як відомо, рівняння регресії є лінійними. Припущення про нормальний розподіл є цілком природним і в ряді випадків може бути обґрунтованим за допомогою граничних теорем теорії імовірностей.

В інших випадках самі величини Y або X можуть не мати нормального розподілу, але деякі функції від них розподілені нормально. Наприклад, відомо, що логарифм доходів населення - нормально розподілена випадкова величина. Цілком природно, наприклад, вважати нормально розподіленою випадковою величиною пробіг автомобіля. Часто гіпотезу про нормальний розподіл приймають в багатьох випадках, коли немає явного їй протиріччя, і, як показує практика, подібна передумова виявляється цілком розумною.

Друга причина, за якою лінійна регресійна модель виявляється переважніше інших, - це менший ризик значної помилки прогнозу. Лінійна регресія знаходить широке застосування в економетриці через чітку економічну інтерпретацію її параметрів.

Лінійна регресія збігається до знаходження рівняння виду

$$\hat{y}_x = a + bx \text{ або } y = a + bx + \varepsilon. \quad (2.6)$$

Рівняння виду $\hat{y}_x = a + bx$ дозволяє за заданим значенням фактору x знаходити теоретичні значення результативної ознаки, підставляючи в нього фактичні значення фактора x .

Побудова лінійної регресії збігається до оцінки її параметрів - a і b . Класичний підхід до оцінювання параметрів лінійної регресії заснований на методі найменших квадратів (МНК), що відповідно до теореми Гаусса-Маркова дає найкращі оцінки цих параметрів.

Теорема Гаусса-Маркова. Якщо регресійна модель (2.6) має наступні властивості:

- збурення ε (або залежна змінна y_i) є величина випадкова, а пояснююча змінна x_i - величина не випадкова;
- математичне сподівання збурення ε дорівнює нулю;

- дисперсія збурення \mathcal{E} (або залежної змінної y_i) постійна для будь-якого i (або $D(y_i) = \sigma^2$) - умова гомоскедастичності або рівномірності збурення (залежної змінної);

- збурення \mathcal{E}_i і \mathcal{E}_j (або змінні y_i і y_j) не корельовані;

- збурення \mathcal{E}_i (або залежна змінна y_i) є нормально розподіленою випадковою величиною,

то оцінки параметрів лінійної регресії мають найменшу дисперсію в класі всіх лінійних незміщених оцінок.

Таким чином, оцінки параметрів лінійної регресії є **ефективними** оцінками.

МНК дозволяє одержати такі оцінки параметрів a і b , при яких сума квадратів відхилень фактичних значень результативної ознаки y від теоретичних \hat{y}_x мінімальна:

$$\sum_{i=1}^n (y_i - \hat{y}_{xi})^2 = \sum_{i=1}^n \mathcal{E}_i^2 \rightarrow \min. \quad (2.7)$$

Тобто з усієї множини ліній лінія регресії на графіку вибирається так, щоб сума квадратів відстаней за вертикаллю між статистичними точками і цією лінією була б мінімальною (рис. 2.2):

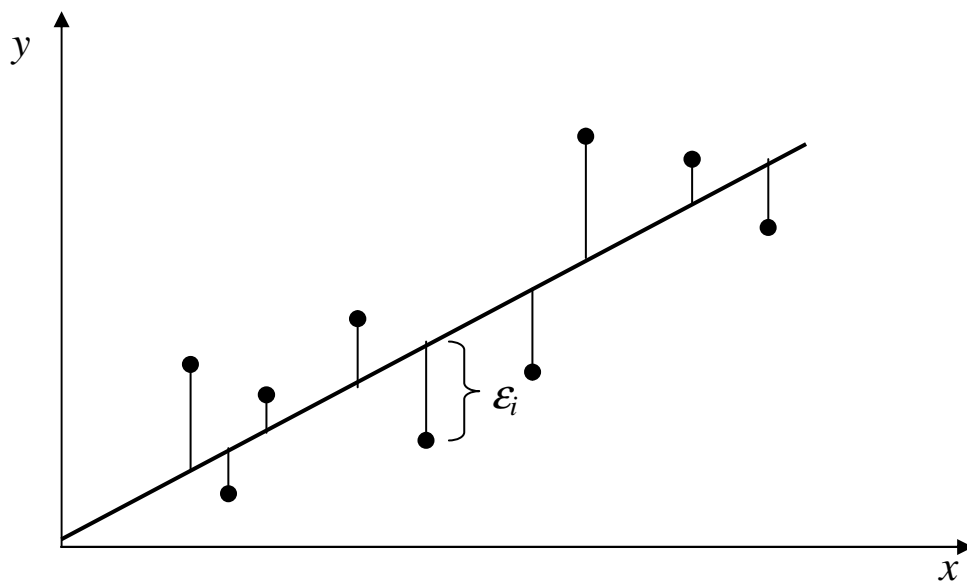


Рисунок 2.2 - Лінія регресії з мінімальною дисперсією залишків

Як відомо з курсу математичного аналізу, щоб знайти мінімум функції (2.7), треба обчислити часткові похідні за кожним з параметрів a і b , дорівнявши їх до нуля.

Позначимо $\sum_{i=1}^n \mathcal{E}_i^2$ через $S(a, b)$, тоді:

$$S(a, b) = \sum_{i=1}^n (y_i - a - bx)^2.$$

$$\begin{cases} \frac{\partial S}{\partial a} = -2 \sum_{i=1}^n (y - a - bx) = 0; \\ \frac{\partial S}{\partial b} = -2 \sum_{i=1}^n (y - a - bx)x = 0. \end{cases} \quad (2.8)$$

Після нескладних перетворень, одержимо наступну систему лінійних рівнянь для оцінки параметрів a і b :

$$\begin{cases} b * \sum x_i^2 + a * \sum x_i = \sum x_i y_i \\ b * \sum x_i + na = \sum y_i \end{cases}. \quad (2.9)$$

Вирішуючи систему рівнянь (2.9), знайдемо шукані оцінки параметрів a і b . Можна скористатися наступними готовими формулами, які впливають безпосередньо з розв'язання системи (2.9):

$$a = \bar{y} - b\bar{x}, \quad b = \frac{\text{cov}(x, y)}{\sigma_x^2}, \quad (2.10)$$

де $\text{cov}(x, y) = \overline{yx} - \bar{y} * \bar{x}$ – коваріація ознак x і y ;

$\sigma_x^2 = \overline{x^2} - \bar{x}^2$ – дисперсія ознаки x

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x, \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y, \quad \overline{yx} = \frac{1}{n} \sum_{i=1}^n yx, \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x^2.$$

Коваріація - числова характеристика спільного розподілу двох випадкових величин, яка дорівнює математичному сподіванню добутку відхилень цих випадкових величин від їхніх математичних сподівань. Дисперсія - характеристика випадкової величини, що визначається як математичне сподівання квадрата відхилення випадкової величини від її математичного сподівання. Математичне сподівання - сума добутків значень випадкової величини та відповідних ймовірностей.

Параметр b називають коефіцієнтом регресії. Його величина показує середню зміну результату зі зміною фактора на одну одиницю.

Можливість чіткої економічної інтерпретації коефіцієнта регресії зробила лінійне рівняння регресії досить розповсюдженим в економетричних дослідженнях.

Формально a – значення y при $x=0$. Якщо фактор-ознака x не може мати нульового значення, то вищевказане трактування вільного члена a не має сенсу, тобто параметр a може не мати економічного смислу.

Рівняння регресії завжди доповнюється показником тісноти зв'язку результативної ознаки Y і фактора-ознаки X . При використанні лінійної регресії як такий показник виступає лінійний коефіцієнт кореляції r_{xy} , який можна розрахувати за наступними формулами:

$$r_{xy} = b \frac{\sigma_x}{\sigma_y} = \frac{\text{cov}(x, y)}{\sigma_x \sigma_y}. \quad (2.11)$$

Лінійний коефіцієнт кореляції знаходиться в межах: $-1 \leq r_{xy} \leq 1$. Чим ближче абсолютне значення r_{xy} до одиниці, тим сильніше лінійний зв'язок між факторами (при $r_{xy} = \pm 1$ має місце строго функціональна залежність). Але необхідно мати на увазі, що близькість абсолютної величини лінійного коефіцієнта кореляції до нуля ще не означає відсутності зв'язку між ознаками. При іншій (нелінійній) специфікації моделі зв'язок між ознаками може виявитися досить тісним.

Для оцінки якості підбору лінійної функції розраховується квадрат лінійного коефіцієнта кореляції r_{xy}^2 , який називають коефіцієнтом детермінації. **Коефіцієнт детермінації** характеризує частку дисперсії результативної ознаки y , що пояснюється регресією, в загальній дисперсії результативної ознаки:

$$r_{xy}^2 = 1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2}, \quad (2.12)$$

$$\text{де } \sigma_{\text{зал}}^2 = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_x)^2, \quad \sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y - \bar{y})^2 = \overline{y^2} - \bar{y}^2.$$

Відповідно величина $1 - r_{xy}^2$ характеризує частку дисперсії y , викликану впливом інших, не врахованих в моделі, факторів.

Після того як рівняння лінійної регресії знайдене, проводять оцінку значущості як рівняння в цілому, так і окремих його параметрів.

2.3 Оцінка значущості рівняння лінійної регресії

Перевірити значущість рівняння регресії означає встановити, чи відповідає математична модель, що виражає залежність між змінними, експериментальним даним і чи досить включених до рівняння пояснюючих змінних (однієї або кількох) для опису залежної змінної.

Щоб мати загальне уявлення про якість моделі, визначають середню помилку апроксимації:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y - \hat{y}_x}{y} \right| * 100\%. \quad (2.13)$$

Середня помилка апроксимації має не перевищувати 8-10%.

Оцінка значущості рівняння регресії в цілому здійснюється за F-критерієм Фішера, розрахунку якого передують дисперсійний аналіз. В математичній статистиці дисперсійний аналіз розглядають як самостійний інструмент статистичного аналізу. В економетриці його застосовують як допоміжний засіб для вивчення якості регресійної моделі.

Відповідно до основної ідеї дисперсійного аналізу, загальну суму квадратів відхилень змінної y від середнього значення \bar{y} розкладають на дві частини – пояснену й непояснену:

$$\sum_{i=1}^n (y - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_x - \bar{y})^2 + \sum_{i=1}^n (y - \hat{y}_x)^2, \quad (2.14)$$

де $\sum_{i=1}^n (y - \bar{y})^2$ – загальна сума квадратів відхилень;

$\sum_{i=1}^n (\hat{y}_x - \bar{y})^2$ – сума квадратів відхилень, пояснена регресією (або факторна сума квадратів відхилень);

$\sum_{i=1}^n (y - \hat{y}_x)^2$ – залишкова сума квадратів відхилень, що характеризує вплив неврахованих у моделі факторів.

Схема дисперсійного аналізу має вигляд, наданий таблицею 2.1 (n - число спостережень, m - число параметрів при змінній x).

Таблиця 2.1 - Схема дисперсійного аналізу

Компоненти дисперсії	Сума квадратів	Число ступенів свободи	Дисперсія на один ступінь свободи
Загальна	$\sum_{i=1}^n (y - \bar{y})^2$	$n-1$	$S_{\text{обиц}}^2 = \frac{\sum_{i=1}^n (y - \bar{y})^2}{n-1}$
Факторна	$\sum_{i=1}^n (\hat{y}_x - \bar{y})^2$	m	$S_{\text{факт}}^2 = \frac{\sum_{i=1}^n (\hat{y}_x - \bar{y})^2}{m}$
Залишкова	$\sum_{i=1}^n (y - \hat{y}_x)^2$	$n-m-1$	$S_{\text{зал}}^2 = \frac{\sum_{i=1}^n (y - \hat{y}_x)^2}{n-m-1}$

Визначення дисперсії на один ступінь свободи призводить дисперсії до порівняного виду. Зіставляючи факторну й залишкову дисперсії, що розраховані на один ступінь свободи, одержимо величину F -критерію Фішера:

$$F = \frac{S_{\text{факт}}^2}{S_{\text{зал}}^2}. \quad (2.15)$$

Фактичне значення F -критерію Фішера (2.15) порівнюють з табличним значенням $F_{\text{табл}}(\alpha, k_1, k_2)$ при рівні значущості α і ступенях свободи $k_1=m$ і $k_2=n-m-1$. При цьому якщо фактичне значення F -критерію перевищує табличне, то признають статистичну значущість рівняння в цілому.

Для парної лінійної регресії $m=1$, тому

$$F = \frac{S_{\text{факт}}^2}{S_{\text{зал}}^2} = \frac{\sum (\hat{y}_x - \bar{y})^2}{\sum (y - \hat{y}_x)^2} * (n-2). \quad (2.16)$$

Величина F -критерію пов'язана з коефіцієнтом детермінації r_{xy}^2 , і її можна розрахувати за наступною формулою:

$$F = \frac{r_{xy}^2}{1 - r_{xy}^2} * (n - 2). \quad (2.17)$$

В парній лінійній регресії оцінюють значущість не тільки рівняння в цілому, але й окремих його параметрів. З цією метою для кожного з параметрів визначають його стандартну помилку: m_a і m_b .

Стандартну помилку коефіцієнта регресії визначають за формулою:

$$m_b = \sqrt{\frac{S_{\text{зал}}^2}{\sum (x - \bar{x})^2}} = \frac{S_{\text{зал}}}{\sigma_x * \sqrt{n}}, \quad (2.18)$$

де $S_{\text{зал}}^2 = \frac{\sum (y - \hat{y}_x)^2}{n - 2}$ – залишкова дисперсія на один ступінь свободи.

Величину стандартної помилки разом з t -розподілом Стюдента при $n-2$ ступенях свободи застосовують для перевірки істотності коефіцієнту регресії і для розрахунку його довірчого інтервалу.

Для оцінки істотності коефіцієнту регресії його величину порівнюють з його стандартною помилкою, тобто визначають фактичне значення t -критерію

Стюдента: $t_b = \frac{b}{m_b}$, яке потім порівнюють з табличним значенням за певним

рівнем значущості α і числом ступенів свободи $n-2$. Довірчий інтервал для коефіцієнта регресії визначають як $b \pm t_{\text{табл}} m_b$. Оскільки знак коефіцієнта регресії вказує на зростання результативної ознаки y при збільшенні ознаки-фактора x ($b > 0$), зменшення результативної ознаки при збільшенні ознаки-фактора ($b < 0$) або його незалежність від незалежної змінної ($b = 0$), то межі довірчого інтервалу для коефіцієнта регресії мають не містити суперечливих результатів (наприклад, $-1,5 \leq b \leq 0,8$). Такого роду запис вказує, що дійсне значення коефіцієнта регресії одночасно містить додатні і від'ємні величини й навіть нуль, чого не може бути.

Стандартну помилку параметра a визначають за формулою:

$$m_a = \sqrt{S_{\text{зал}}^2 \frac{\sum x^2}{n \sum (x - \bar{x})^2}} = S_{\text{зал}} \frac{\sqrt{\sum x^2}}{\sigma_x n}. \quad (2.19)$$

Процедура оцінювання істотності даного параметра не відрізняється від розглянутої вище для коефіцієнта регресії. Обчислюють t -критерій: $t_a = \frac{a}{m_a}$,

його величину порівнюють з табличним значенням при $n-2$ ступенях свободи.

Значущість лінійного коефіцієнта кореляції перевіряють на основі величини помилки коефіцієнта кореляції m_r :

$$m_r = \sqrt{\frac{1 - r^2}{n - 2}}. \quad (2.20)$$

Фактичне значення t -критерію Стюдента визначається як $t_r = \frac{r}{m_r}$.

Існує зв'язок між t -критерієм Стюдента і F -критерієм Фішера:

$$t_b = t_r = \sqrt{F} . \quad (2.21)$$

В прогнозах розрахунках за рівнянням регресії визначають прогнозоване \hat{y}_p значення як точковий прогноз \hat{y}_x при $x_p = x_k$, тобто шляхом підстановки до рівняння регресії $\hat{y}_x = a + bx$ відповідного значення x . Однак точковий прогноз явно не реальний. Тому він доповнюється розрахунком стандартної помилки \hat{y}_p , тобто $M(\hat{y}_p)$, і відповідно інтервальною оцінкою прогнозного значення \hat{y}_p :

$$\hat{y}_p - \Delta_{\hat{y}_p} \leq \hat{y}_p \leq \hat{y}_p + \Delta_{\hat{y}_p} ,$$

де $\Delta_{\hat{y}_p} = m_{\hat{y}_p} * t_{табл}$, а $m_{\hat{y}_p}$ – середня помилка прогнозованого індивідуального значення:

$$m_{\hat{y}_p} = S_{зал} \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{n\sigma_x^2}} . \quad (2.22)$$

Запитання для самоперевірки

1. Поясніть терміни «рівняння регресії» і «функція регресії».
2. Які помилки належать до помилок специфікації? До помилок виміру?
3. На яких підставах обирають вид регресійної залежності?
4. Як розраховують залишкову дисперсію? Загальну дисперсію ознаки y ?
5. У чому полягає сутність методу найменших квадратів?
6. Якими властивостями має володіти лінійна модель, щоб оцінки її параметрів мали найменшу дисперсію в класі всіх лінійних незміщених оцінок?
7. Поясніть, що таке коефіцієнт регресії? Коефіцієнт кореляції? Коефіцієнт детермінації?
8. Як визначають статистичну значущість рівняння регресії в цілому?
9. З якою метою визначають стандартну помилку коефіцієнта регресії? Який критерій для цього використовують?
10. Як визначають довірчий інтервал для коефіцієнта регресії?
11. З якою метою і як визначають значущість лінійного коефіцієнта кореляції?
12. У якому випадку використовують лінійну модель множинної регресії?
13. Що таке коефіцієнти «чистої» регресії? Який в них економічний зміст?
14. У якому випадку використовують стандартизовані коефіцієнти регресії?
15. Що характеризують показник множинної кореляції та показник детермінації?

16. Що таке лінійний коефіцієнт множинної кореляції і скорегований індекс множинної кореляції?
17. Для чого розраховують часткові коефіцієнти кореляції?
18. Як оцінюють значущість рівняння множинної регресії в цілому та значущість коефіцієнтів чистої регресії?
19. Поясніть, у чому полягає процедура відбору факторів під час побудови рівняння регресії за методом виключення?
20. Охарактеризуйте помилки специфікації моделі. У чому вони полягають?
21. Поясніть зміст коефіцієнта регресії, назвіть способи його оцінювання, покажіть, як його використовують для розрахунку мультиплікатора у функції споживання.
22. Поясніть, що розуміють під кількістю ступенів свободи. Як її визначають для факторної та залишкової сум квадратів?
23. Поясніть призначення та концепцію критерію Фішера.
24. Як оцінюють значущість параметрів рівняння регресії?
25. На яких посиленнях базується використання методу найменших квадратів?
26. Які властивості оцінюваних параметрів забезпечує використання методу найменших квадратів?
27. Поясніть сутність середньої помилки апроксимації. Як її визначають?
28. Що оцінюють за допомогою коефіцієнта еластичності?

Тема 3

МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ

3.1 Лінійна модель множинної регресії

Економічні явища, як правило, визначаються великим числом одночасно і сукупно діючих факторів. У зв'язку з цим часто виникає задача дослідження залежності однієї залежної змінної Y від кількох пояснюючих змінних X_1, X_2, \dots, X_n . Цю задачу вирішують за допомогою множинного регресійного аналізу. Парна регресія може дати добрий результат при моделюванні, якщо впливом інших факторів можна зневажити. Якщо ж цим впливом зневажити не можна, то слід спробувати виявити вплив інших факторів, увівши їх до моделі, тобто побудувати рівняння множинної регресії

$$\hat{y} = f(x_1, x_2, \dots, x_m), \quad (3.1)$$

де y – залежна змінна (результативна ознака);

x_i – незалежні, або пояснюючі, змінні (ознаки-фактори).

Множинну регресію широко використовують у розв'язанні проблем попиту, прибутковості акцій, при вивченні функції витрат виробництва, у макроекономічних розрахунках і цілому ряді інших питань економетрики. В цей час множинна регресія – один з найпоширеніших методів в економетриці. Основна мета множинної регресії – побудувати модель з великим числом факторів, визначивши при цьому вплив кожного з них окремо, а також сукупний їхній вплив на показник, що моделюється. Можливі різні види рівнянь множинної регресії – лінійні та нелінійні.

У зв'язку з чіткою інтерпретацією параметрів найбільш широко використовують лінійну функцію. У лінійній множинній регресії $\hat{y}_x = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$ параметри при x називаються коефіцієнтами «чистої» регресії. Вони характеризують середню зміну результативної ознаки зі зміною відповідного фактору на одиницю при незмінному значенні інших факторів, зафіксованих на середньому рівні.

Розглянемо лінійну модель множинної регресії

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon. \quad (3.2)$$

Класичний підхід до оцінювання параметрів лінійної моделі множинної регресії заснований на методі найменших квадратів (МНК). Теорема Гаусса-Маркова, що розглянута для парної регресійної моделі, виявляється вірною і для моделі множинної регресії. Тобто МНК дає найбільш ефективні (тобто такі, що мають найменшу дисперсію в класі лінійних незміщених оцінок) оцінки параметрів регресійної моделі.

Нагадаємо, що МНК дозволяє одержати такі оцінки параметрів, за якими сума квадратів відхилень фактичних значень результативної ознаки у від розрахункових \hat{y}_x мінімальна:

$$\sum_i (y_i - \hat{y}_{x_i})^2 \rightarrow \min. \quad (3.3)$$

З виразу (3.3) дістанемо функцію $m+1$ аргументу:

$$S(a,b_1,b_2,...,b_m)=\sum_{i=1}^n (y-a-b_1x_1-b_2x_2-...-b_mx_m)^2.$$

Визначивши часткові похідні першого порядку, дорівнюємо їх до нуля:

$$\left\{ \begin{aligned} \frac{\partial S}{\partial a} &= -2 \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) = 0, \\ \frac{\partial S}{\partial b_1} &= -2 \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) x_1 = 0, \\ &\dots \dots \dots \\ \frac{\partial S}{\partial b_m} &= -2 \sum_{i=1}^n (y - a - b_1 x_1 - b_2 x_2 - \dots - b_m x_m) x_m = 0. \end{aligned} \right.$$

Після елементарних перетворень приходимо до системи лінійних нормальних рівнянь для знаходження параметрів лінійного рівняння множинної регресії (3.2):

[illegible]

Для двофакторної моделі дана система матиме вигляд:

$$\begin{cases} na + b_1 \sum_{i=1}^n x_{i1} + b_2 \sum_{i=1}^n x_{i2} = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + b_2 \sum_{i=1}^n x_{i1} x_{i2} = \sum_{i=1}^n x_{i1} y_i, \\ a \sum_{i=1}^n x_{i2} + b_1 \sum_{i=1}^n x_{i1} x_{i2} + b_2 \sum_{i=1}^n x_{i2}^2 = \sum_{i=1}^n x_{i2} y_i. \end{cases} \quad (3.5)$$

На практиці часто виникає необхідність порівняння впливу на результативну ознаку різних пояснюючих змінних. Для цього необхідно, щоб ці пояснюючі змінні були приведені до однакових одиниць виміру. В таких випадках використовують стандартизовані коефіцієнти регресії:

$$t_y = \frac{y - \bar{y}}{\sigma_y}. \quad (3.6)$$

Метод найменших квадратів можна застосувати і до рівняння множинної регресії в стандартизованому виді:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_m t_{x_m} + \varepsilon, \quad (3.7)$$

де $t_y, t_{x_1}, \dots, t_{x_m}$ – стандартизовані змінні: $t_y = \frac{y - \bar{y}}{\sigma_y}$, $t_{x_i} = \frac{x_i - \bar{x_i}}{\sigma_{x_i}}$, для яких

середнє значення дорівнює нулю: $\bar{t}_y = \bar{t}_{x_i} = 0$, а середнє квадратичне відхилення дорівнює одиниці: $\sigma_{t_y} = \sigma_{t_{x_i}} = 1$;

β_i – стандартизовані коефіцієнти регресії.

Стандартизовані коефіцієнти регресії показують, на скільки одиниць зміниться в середньому результат, якщо відповідний фактор x_i зміниться на одну одиницю при незмінному середньому рівні інших факторів. В силу того, що всі змінні задані як центровані й нормовані, стандартизовані коефіцієнти регресії β_i можна порівнювати один з одним. Порівнюючи їх один з одним, можна ранжувати фактори за силою їхнього впливу на результативну ознаку. В цьому основна перевага стандартизованих коефіцієнтів регресії на відміну від коефіцієнтів «чистої» регресії, які непорівнянні один з одним.

Застосовуючи МНК до рівняння множинної регресії в стандартизованому виді, одержимо систему нормальних рівнянь

[illegible]

де r_{yx_i} й $r_{x_i x_i}$ – коефіцієнти парної і міжфакторної кореляції.

Коефіцієнти «чистої» регресії b_i пов'язані із стандартизованими коефіцієнтами регресії β_i у такий спосіб:

$$b_i = \beta_i \frac{\sigma_y}{\sigma_{x_i}}. \quad (3.9)$$

Тому можна переходити від рівняння регресії в стандартизованому масштабі (3.7) до рівняння регресії в натуральному масштабі змінних (3.2), при цьому параметр a визначається як $a = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2 - \dots - b_m \bar{x}_m$.

Зміст стандартизованих коефіцієнтів регресії дозволяє використовувати їх при відсіванні факторів – з моделі виключають фактори з найменшим значенням β_i .

На основі лінійного рівняння множинної регресії (3.2) можна знайти часткові рівняння регресії:

$$\left\{ \begin{array}{l} y_{x_1 x_2, x_3, \dots, x_m} = \hat{f}(x_1), \\ y_{x_2 x_1, x_3, \dots, x_m} = \hat{f}(x_2), \\ \\ y_{x_m x_1, x_2, \dots, x_{m-1}} = \hat{f}(x_m), \end{array} \right. \quad (3.10)$$

тобто рівняння регресії, які зв'язують результативну ознаку з відповідним фактором x_i при закріпленні інших факторів на середньому рівні. В розгорнутому вигляді систему (3.10) можна переписати наступним чином:

[illegible]

При підстановці в ці рівняння середніх значень відповідних факторів вони приймають вигляд парних рівнянь лінійної регресії:

[illegible]

де

вих рівнянь регресії визначати часткові коефіцієнти еластичності:

$$E_{yx_i} = b_i \frac{x_i}{\hat{y}_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m}}, \quad (3.12)$$

де b_i – коефіцієнт регресії для фактора x_i у рівнянні множинної регресії;

$$\hat{y}_{x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_m} - \text{часткове рівняння регресії.}$$

Поряд з частковими коефіцієнтами еластичності можна знайти середні за сукупністю показники еластичності:

$$\overline{E}_i = b_i \frac{\overline{x}_i}{\overline{y}_{x_i}}, \quad (3.13)$$

які показують, на скільки відсотків у середньому зміниться результативна ознака при зміні відповідного фактора на 1%. Середні показники еластичності можна порівнювати один з одним і відповідно ранжувати фактори за силою їхнього впливу на результативну ознаку.

3.2 Оцінка значущості множинної регресії і показники якості моделі

Практичну значущість рівняння множинної регресії оцінюють за допомогою показника множинної кореляції і його квадрата - показника детермінації.

Показник множинної кореляції характеризує тісноту зв'язку розглянутого набору факторів з досліджуваною ознакою або, інакше, оцінює тісноту загального впливу факторів на результативну ознаку.

Незалежно від форми зв'язку показник множинної кореляції може бути знайдений як індекс множинної кореляції:

$$R_{y x_1 x_2 \dots x_m} = \sqrt{1 - \frac{\sigma_{sal}^2}{\sigma_y^2}}, \quad (3.14)$$

де σ_y^2 – загальна дисперсія результативної ознаки;

$\sigma_{\text{зал}}^2$ – залишкова дисперсія.

Межі зміни індексу множинної кореляції - від 0 до 1. Чим ближче його значення до 1, тим тісніше зв'язок результативної ознаки з усім набором досліджуваних факторів. Величина індексу множинної кореляції повинна бути біль-

ше (або дорівнювати) максимального значення парного індексу кореляції:

$$R_{yx_1x_2\dots x_m} \geq r_{yx_i(\max)}, \quad i = \overline{1, m}.$$

При правильному включенні факторів до регресійної моделі величина індексу множинної кореляції буде істотно відрізнятися від індексу кореляції парної залежності. Якщо ж додатково включені до рівняння множинної регресії фактори третьорядні, то індекс множинної кореляції може практично збігатися з індексом парної кореляції (розходження в третьому, четвертому знаках). Звідки випливає, що порівнюючи індекси множинної і парної кореляції, можна зробити висновок про доцільність включення до рівняння регресії того або іншого фактора.

Розрахунок індексу множинної кореляції передбачає визначення залишкової дисперсії:

$$\sigma_{\text{зал}}^2 = \frac{1}{n} \sum (y - \hat{y}_{x_1x_2\dots x_m})^2. \quad (3.15)$$

Можна користуватися формулою для індексу множинної детермінації:

$$R_{yx_1x_2\dots x_m}^2 = 1 - \frac{\sum (y - \hat{y}_{x_1x_2\dots x_m})^2}{\sum (y - \bar{y})^2}. \quad (3.16)$$

При лінійній залежності ознак формула для індексу множинної кореляції може бути представлена в такий спосіб:

$$R_{yx_1x_2\dots x_m} = \sqrt{\sum_{i=1}^n \beta_i r_{yx_i}}, \quad (3.17)$$

де β_i – стандартизовані коефіцієнти регресії;

r_{yx_i} – парні коефіцієнти кореляції результативної ознаки з кожним з факторів.

Вираз індексу множинної кореляції для лінійної регресії називають **лінійним коефіцієнтом множинної кореляції**, або **сукупним коефіцієнтом кореляції**.

Сукупний коефіцієнт кореляції можна також визначити за допомогою матриці парних коефіцієнтів кореляції:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - \frac{\Delta r}{\Delta r_{11}}}, \quad (3.18)$$

де

$$\Delta r = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_p} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_p} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_p} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_p} & r_{x_px_1} & r_{x_px_2} & \dots & 1 \end{vmatrix}$$

- визначник матриці парних коефіцієнтів кореляції;

$$\Delta r_{11} = \begin{vmatrix} 1 & r_{x_1 x_2} & \dots & r_{x_1 x_p} \\ r_{x_2 x_1} & 1 & \dots & r_{x_2 x_p} \\ \dots & \dots & \dots & \dots \\ r_{x_p x_1} & r_{x_p x_2} & \dots & 1 \end{vmatrix} \quad - \text{ визначник матриці міжфакторної кореляції.}$$

Як бачимо, величина множинного коефіцієнта кореляції залежить не тільки від кореляції результативної ознаки з кожним з факторів, але й від міжфакторної кореляції. Таким чином, визначити сукупний коефіцієнт кореляції можна, не звертаючись до рівняння множинної регресії, а використовуючи лише парні коефіцієнти кореляції.

У розглянутих показниках множинної кореляції (індекс і коефіцієнт) використовують значення залишкової дисперсії, що має систематичну помилку вбік зменшення. В результаті залишкова дисперсія тим менша, чим більше параметрів визначають у рівнянні регресії при заданому обсязі спостережень n . Якщо число параметрів при x_i дорівнює m і наближається до обсягу спостережень, то залишкова дисперсія буде близькою до нуля, і коефіцієнт (індекс) кореляції наблизиться до одиниці навіть при слабкому зв'язку факторів з результативною ознакою. Для усунення цього явища використовують скорегований індекс (коефіцієнт) множинної кореляції.

Скорегований індекс множинної кореляції містить виправлення на число ступенів свободи, а саме залишкова сума квадратів $\sum (y - \hat{y}_{x_1, x_2, \dots, x_m})^2$ ділиться на число ступенів свободи залишкової варіації $n-m-1$, а загальна сума квадратів відхилень $\sum (y - \bar{y})^2$ - на число ступенів свободи в цілому за сукупністю $(n-1)$.

Нагадаємо, що R^2 характеризує частку варіації залежної змінної, яка зумовлена регресією або мінливістю пояснюючих змінних. Чим ближче R^2 до одиниці, тим краще регресія описує залежність між пояснюючими змінними і результативною ознакою. Разом з тим використання тільки одного коефіцієнта детермінації R^2 для вибору найкращого рівняння регресії може виявитися недостатнім. На практиці зустрічаються випадки, коли погано визначена модель регресії може дати порівняно високий коефіцієнт R^2 .

Недоліком коефіцієнта детермінації R^2 також є те, що він збільшується при додаванні нових пояснюючих змінних у модель, хоча це й не обов'язково поліпшує якість регресійної моделі. В цьому сенсі переважніше використовувати скорегований (адаптований, поправлений) коефіцієнт детермінації \hat{R}^2 , визначений за формулою

$$\hat{R}^2 = 1 - \frac{n-1}{n-p-1} (1 - R^2). \quad (3.19)$$

З (3.19) випливає, що чим більше число пояснюючих змінних p , тим менше \hat{R}^2 порівняно з R^2 . На відміну від R^2 скорегований коефіцієнт \hat{R}^2 може зменшуватися при введенні до моделі нових пояснюючих змінних, які не роблять істотного впливу на залежну змінну. Однак навіть збільшення скорегованого коефіцієнта детермінації \hat{R}^2 при введенні до моделі нової пояснюючої змінної не завжди означає, що її коефіцієнт регресії є значущим (це відбувається, як можна показати, тільки у випадку, якщо відповідне значення t -статистики перевищує одиницю (за абсолютною величиною), тобто $|t| > 1$. Інакше кажучи, збільшення \hat{R}^2 ще не означає поліпшення якості регресійної моделі.

Як було показано вище, ранжирування факторів, що беруть участь у множинній лінійній регресії, може бути проведене через стандартизовані коефіцієнти регресії (β -коефіцієнти). Ця сама мета може бути досягнута за допомогою часткових коефіцієнтів кореляції (для лінійних зв'язків). Окрім того, часткові показники кореляції широко використовують при розв'язанні проблеми відбору факторів: доцільність включення того або іншого фактора до моделі можна довести величиною показника часткової кореляції.

Часткові коефіцієнти кореляції характеризують тісноту зв'язку між результативною ознакою й відповідним фактором при елімінуванні (усуненні впливу) інших факторів, включених до рівняння регресії.

Показники часткової кореляції являють собою відношення зменшення залишкової дисперсії за рахунок додаткового включення до аналізу нового фактора до залишкової дисперсії, що мала місце до введення його у модель.

В загальному вигляді при наявності m факторів для рівняння

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + \varepsilon$$

коефіцієнт часткової кореляції, що вимірює вплив на y фактору x_i , при незмінному рівні інших факторів, можна визначити за формулою:

$$r_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m} = \sqrt{1 - \frac{1 - R_{yx_1 x_2 \dots x_i \dots x_m}^2}{1 - R_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m}^2}}, \quad (3.20)$$

де $R_{yx_1 x_2 \dots x_i \dots x_m}^2$ – множинний коефіцієнт детермінації всіх m факторів з результативною ознакою;

$R_{yx_1 x_2 \dots x_{i-1} x_{i+1} \dots x_m}^2$ – той самий показник детермінації, але без введення до моделі фактору x_i .

При двох факторах формула (3.20) прийме вигляд:

$$r_{yx_1 x_2} = \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - R_{yx_2}^2}}, \quad r_{yx_2 x_1} = \sqrt{1 - \frac{1 - R_{yx_1 x_2}^2}{1 - R_{yx_1}^2}}. \quad (3.21)$$

Порядок часткового коефіцієнта кореляції визначають кількістю факторів, вплив яких виключається. Наприклад, $r_{yx_1x_2}$ – коефіцієнт часткової кореляції першого порядку. Відповідно коефіцієнти парної кореляції називають коефіцієнтами нульового порядку. Коефіцієнти часткової кореляції більш високих порядків можна визначити через коефіцієнти часткової кореляції більш низьких порядків за рекурентною формулою:

$$r_{yx_1x_2 \dots x_{i-1}x_{i+1} \dots x_m} = \frac{r_{yx_1x_2 \dots x_{i-1}x_{i+1} \dots x_{m-1}} - r_{yx_m x_1x_2 \dots x_{m-1}} * r_{yx_i x_m x_1x_2 \dots x_{i-1}x_{i+1} \dots x_{m-1}}}{\sqrt{(1 - r_{yx_m x_1x_2 \dots x_{m-1}}^2) * (1 - r_{yx_i x_m x_1x_2 \dots x_{i-1}x_{i+1} \dots x_{m-1}}^2)}}. \quad (3.22)$$

За двох факторів дана формула прийме вигляд:

$$r_{yx_1x_2} = \frac{r_{yx_1} - r_{yx_2} * r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2) * (1 - r_{x_1x_2}^2)}}; \quad r_{yx_2x_1} = \frac{r_{yx_2} - r_{yx_1} * r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2) * (1 - r_{x_1x_2}^2)}}. \quad (3.23)$$

Для рівняння регресії з трьома факторами часткові коефіцієнти кореляції другого порядку визначають на основі часткових коефіцієнтів кореляції першого порядку. Так, за рівнянням $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \varepsilon$ можливе вираховування трьох часткових коефіцієнтів кореляції другого порядку:

$$r_{yx_1x_2x_3}, \quad r_{yx_2x_1x_3}, \quad r_{yx_3x_1x_2},$$

кожний з яких визначають за рекурентною формулою. Наприклад, при $i=1$ маємо формулу для розрахунку $r_{yx_1x_2x_3}$:

$$r_{yx_1x_2x_3} = \frac{r_{yx_1x_2} - r_{yx_3x_2} * r_{x_1x_2x_3}}{\sqrt{(1 - r_{yx_3x_2}^2) * (1 - r_{x_1x_2x_3}^2)}}. \quad (3.24)$$

Розраховані за рекурентною формулою часткові коефіцієнти кореляції змінюються в межах від -1 до +1, а за формулами через множинні коефіцієнти детермінації – від 0 до 1. Порівняння їх одного з одним дозволяє ранжувати фактори за тісністю їхнього зв'язку з результативною ознакою. Часткові коефіцієнти кореляції дають міру тісноти зв'язку кожного фактора з результативною ознакою в чистому вигляді. Якщо із стандартизованого рівняння регресії $t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \beta_3 t_{x_3} + \varepsilon$ випливає, що $\beta_1 > \beta_2 > \beta_3$, тобто за силою впливу на результативну ознаку порядок факторів такий: x_1, x_2, x_3 , то цей самий порядок факторів визначають і за співвідношенням часткових коефіцієнтів кореляції,

$$r_{yx_1x_2x_3} > r_{yx_2x_1x_3} > r_{yx_3x_1x_2}.$$

В економетриці часткові коефіцієнти кореляції зазвичай не мають самостійного значення. Їх використовують на стадії формування моделі. Так, будуючи багатфакторну модель, на першому кроці визначають рівняння регресії з повним набором факторів і розраховують матрицю часткових коефіцієнтів кореляції. На другому кроці відбирають фактор з найменшою і неістотною за t -

критерієм Стюдента величиною показника часткової кореляції. Виключивши його з моделі, будують нове рівняння регресії. Процедура триває доти, поки не виявиться, що всі часткові коефіцієнти кореляції істотно відрізняються від нуля. Якщо виключено неістотний фактор, то множинні коефіцієнти детермінації на двох суміжних кроках побудови регресійної моделі майже не відрізняються один від одного, $R_{m+1}^2 \approx R_m^2$, де m – число факторів.

З наведених вище формул часткових коефіцієнтів кореляції видний зв'язок цих показників із сукупним коефіцієнтом кореляції. Знаючи часткові коефіцієнти кореляції (послідовно першого, другого і більш високого порядку), можна визначити сукупний коефіцієнт кореляції за формулою:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - (1 - r_{yx_1}^2) * (1 - r_{yx_2x_1}^2) * (1 - r_{yx_3x_1x_2}^2) * \dots * (1 - r_{yx_mx_1x_2\dots x_{m-1}}^2)}. \quad (3.25)$$

Зокрема, для двофакторного рівняння формула (3.25) приймає вигляд:

$$R_{yx_1x_2\dots x_m} = \sqrt{1 - (1 - r_{yx_1}^2) * (1 - r_{yx_2x_1}^2)}. \quad (3.26)$$

При повній залежності результативної ознаки від досліджуваних факторів коефіцієнт їхнього сукупного впливу дорівнює одиниці. Від одиниці віднімають частку залишкової варіації результативної ознаки $(1 - r^2)$, що зумовлена послідовно включеними до аналізу факторами. В результаті підкореневий вираз характеризує сукупну дію всіх досліджуваних факторів.

Значущість рівняння множинної регресії в цілому, так само як і в парній регресії, оцінюють за допомогою F -критерію Фішера:

$$F = \frac{S_{факт}}{S_{зал}} = \frac{R^2}{1 - R^2} * \frac{n - m - 1}{m}, \quad (3.27)$$

де $S_{факт}$ – факторна сума квадратів на один ступінь свободи;

$S_{зал}$ – залишкова сума квадратів на один ступінь свободи;

R^2 – коефіцієнт (індекс) множинної детермінації;

m – число параметрів при змінних x (в лінійній регресії збігається з числом включених до моделі факторів);

n – число спостережень.

Оцінюють значущість не тільки рівняння в цілому, але й фактора, додатково включеного до регресійної моделі. Необхідність такої оцінки пов'язана з тим, що не кожен фактор, що ввійшов до моделі, може істотно збільшувати частку поясненої варіації результативної ознаки. Окрім того, за наявності в моделі кількох факторів, їх можна вводити до моделі у різній послідовності. Через кореляцію між факторами значущість того самого фактора може бути різною залежно від послідовності його введення до моделі. Мірою для оцінки включення фактора до моделі служить частковий F -критерій, тобто F_{x_i} .

Частковий F -критерій заснований на порівнянні приросту факторної дисперсії, зумовленого впливом додатково включеного фактора, із залишковою дисперсією на один ступінь свободи за регресійною моделлю в цілому. В загальному вигляді для фактора x_i частковий F -критерій визначиться як

$$F_{x_i} = \frac{R^2_{yx_1 \dots x_i \dots x_m} - R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}}{1 - R^2_{yx_1 \dots x_i \dots x_m}} * \frac{n - m - 1}{1}, \quad (3.28)$$

де $R^2_{yx_1 \dots x_i \dots x_m}$ – коефіцієнт множинної детермінації для моделі з повним набором факторів;

$R^2_{yx_1 \dots x_{i-1} x_{i+1} \dots x_m}$ – той самий показник, але без включення до моделі фактора x_i ;

n – число спостережень;

m – число параметрів у моделі (без вільного члена).

Фактичне значення часткового F -критерію порівнюють з табличним за рівнем значущості α і числом ступенів свободи: 1 і $n-m-1$. Якщо фактичне значення F_{x_i} перевищує $F(\alpha, k_1, k_2)$, то додаткове включення фактора x_i у модель статистично виправдане і коефіцієнт чистої регресії b_i при факторі x_i є статистично значущим. Якщо ж фактичне значення F_{x_i} менше табличного, то додаткове включення до моделі фактора x_i не збільшує істотно частку поясненої варіації ознаки y , отже, недоцільно його включення до моделі; коефіцієнт регресії при даному факторі в цьому випадку є статистично не значущим.

Для двофакторного рівняння часткові F -критерії мають вигляд:

$$F_{x_1} = \frac{R^2_{yx_1 x_2} - r^2_{yx_2}}{1 - R^2_{yx_1 x_2}} * (n - 3); \quad F_{x_2} = \frac{R^2_{yx_1 x_2} - r^2_{yx_1}}{1 - R^2_{yx_1 x_2}} * (n - 3). \quad (3.29)$$

За допомогою часткового F -критерію можна перевірити значущість всіх коефіцієнтів регресії у припущенні, що кожний відповідний фактор x_i вводився до рівняння множинної регресії останнім.

Частковий F -критерій оцінює значущість коефіцієнтів чистої регресії. Знаючи величину F_{x_i} , можна визначити і t -критерій для коефіцієнта регресії при i -му факторі, t_{b_i} , а саме:

$$t_{b_i} = \sqrt{F_{x_i}}. \quad (3.30)$$

Оцінка значущості коефіцієнтів чистої регресії за t -критерієм Стюдента може бути проведена і без розрахунку часткових F -критеріїв. В цьому випадку, як і в парній регресії, для кожного фактора використовують формулу:

$$t_{b_i} = \frac{b_i}{m_{b_i}}, \quad (3.31)$$

де b_i – коефіцієнт чистої регресії при факторі x_i ;

m_{b_i} – середня квадратична (стандартна) помилка коефіцієнта регресії b_i .

Для рівняння множинної регресії $\hat{y} = a + b_1x_1 + b_2x_2 + \dots + b_mx_m$ середню квадратичну помилку коефіцієнта регресії можна визначити за наступною формулою:

$$m_{b_1} = \frac{\sigma_y \sqrt{1 - R_{yx_1 \dots x_m}^2}}{\sigma_{x_1} \sqrt{1 - R_{x_1 x_1 \dots x_m}^2}} \frac{1}{\sqrt{n - m - 1}}, \quad (3.32)$$

де σ_y – середнє квадратичне відхилення для ознаки y ;

σ_{x_i} – середнє квадратичне відхилення для ознаки x_i ;

$R_{yx_1 \dots x_m}^2$ – коефіцієнт детермінації для рівняння множинної регресії;

$R_{x_i x_1 \dots x_m}^2$ – коефіцієнт детермінації для залежності фактору x_i з усіма іншими факторами рівняння множинної регресії;

$n-m-1$ – число ступенів свободи для залишкової суми квадратів відхилень.

Як бачимо, для використання даної формули необхідно скористатися матрицю міжфакторної кореляції і розрахувати за нею відповідні коефіцієнти детермінації $R_{x_i x_1 \dots x_m}^2$. Так, для рівняння $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3$ оцінка значущості коефіцієнтів регресії b_1, b_2, b_3 припускає розрахунок трьох міжфакторних коефіцієнтів детермінації: $R_{x_1 x_2 x_3}^2, R_{x_2 x_1 x_3}^2, R_{x_3 x_1 x_2}^2$.

Взаємозв'язок показників часткового коефіцієнта кореляції, часткового F -критерію і t -критерію Стюдента для коефіцієнтів чистої регресії може використовуватися в процедурі відбору факторів. Відсівання факторів при побудові рівняння регресії методом виключення практично можна здійснювати не тільки за частковими коефіцієнтами кореляції, виключаючи на кожному кроці фактор з найменшим незначущим значенням часткового коефіцієнта кореляції, але й за величинами t_{b_i} і F_{x_i} . Частковий F -критерій широко використовують і при побудові моделі за методом включення змінних і за кроковим регресійним методом.

3.3 Мультиколінеарність та її вплив на оцінки параметрів моделі

Однією з проблем, пов'язаних з практичним використанням моделі множинної регресії, є мультиколінеарність.

Під **мультиколінеарністю** розуміють високу взаємну корельованість пояснюючих змінних. Мультиколінеарність може проявлятися у функціональній (явній) і стохастичній (схованій) формах.

При функціональній формі мультиколінеарності принаймні один з парних зв'язків між пояснюючими змінними є лінійною функціональною залежністю. Це призводить до неможливості розв'язання відповідної системи нормальних рівнянь і одержання оцінок параметрів регресійної моделі.

Проте в економічних дослідженнях мультиколінеарність частіше проявляється в стохастичній формі, коли хоча б між двома пояснюючими змінними існує тісний кореляційний зв'язок. Оцінки параметрів стають дуже чутливими до незначної зміни результатів спостережень і обсягу вибірки. Рівняння регресії в цьому випадку, як правило, не мають реального змісту, тому що деякі з коефіцієнтів можуть мати неправильні з погляду економічної теорії знаки і невідповідно великі значення.

Точних кількісних критеріїв для визначення наявності або відсутності мультиколінеарності не існує. Проте, є деякі евристичні підходи до її виявлення. Один з таких підходів полягає в аналізі кореляційної матриці між пояснюючими змінними x_j і виявленні пар змінних, що мають високі коефіцієнти кореляції (зазвичай понад 0,8). Якщо такі змінні існують, то вважають, що між ними має місце мультиколінеарність.

Корисно також знаходити множинні коефіцієнти детермінації між однією з пояснюючих змінних і деякою групою з них. Наявність високого множинного коефіцієнта детермінації (зазвичай понад 0,6) свідчить про мультиколінеарність.

Для усунення або зменшення мультиколінеарності використовують ряд методів. Найпростіший з них (але далеко не завжди можливий) полягає в тому, що з двох пояснюючих змінних, які мають високий коефіцієнт кореляції (понад 0,8), одну змінну виключають з розгляду. При цьому, яку змінну залишити, а яку видалити з аналізу, вирішують у першу чергу на підставі економічних міркувань. Якщо з економічної точки зору ні одній зі змінних не можна надати перевагу, то залишають ту з двох змінних, що має більший коефіцієнт кореляції із залежною змінною.

У цій вимозі проявляється специфіка множинної регресії як методу дослідження комплексного впливу факторів в умовах їхньої незалежності один від одного.

За величиною парних коефіцієнтів кореляції виявляється лише явна колінеарність факторів. Найбільші труднощі у використанні апарата множинної регресії виникають при наявності мультиколінеарності факторів, коли понад два фактори зв'язані один з одним лінійною залежністю, тобто має місце сукупний вплив факторів один на одного. Наявність мультиколінеарності факторів може означати, що деякі фактори будуть завжди діяти в унісон. У результаті варіація у вихідних даних перестає бути повністю незалежною і не можна оцінити вплив кожного фактора окремо.

Включення до моделі мультиколінеарних факторів небажане в силу наступних наслідків:

- утруднюється інтерпретація параметрів множинної регресії як характеристик дії факторів в «чистому» виді, тому що фактори корельовані; параметри лінійної регресії втрачають економічний зміст;
- оцінки параметрів ненадійні, виявляють великі стандартні помилки й змінюються зі зміною обсягу спостережень (не тільки за величиною, але й за знаком), що робить модель непридатною для аналізу й прогнозування.

Для оцінки мультиколінеарності факторів може використовуватися визначник матриці парних коефіцієнтів кореляції між факторами.

Якщо фактори не корельовані один з одним, то матриця парних коефіцієнтів кореляції між факторами є одиничною матрицею, оскільки всі недіагональні елементи $r_{x_i y_j}$ ($i \neq j$) дорівнюють нулю. Так, для рівняння, що включає три пояснюючих змінних,

$$\hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_m,$$

матриця коефіцієнтів кореляції між факторами має визначник, що дорівнює одиниці:

$$\text{Det} = \begin{vmatrix} r_{x_1 x_1} & r_{x_1 x_2} & r_{x_1 x_3} \\ r_{x_2 x_1} & r_{x_2 x_2} & r_{x_2 x_3} \\ r_{x_3 x_1} & r_{x_3 x_2} & r_{x_3 x_3} \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} = 1.$$

Якщо ж між факторами існує повна лінійна залежність, і всі коефіцієнти кореляції дорівнюють одиниці, то визначник такої матриці дорівнює нулю:

$$\text{Det} = \begin{vmatrix} r_{x_1 x_1} & r_{x_1 x_2} & r_{x_1 x_3} \\ r_{x_2 x_1} & r_{x_2 x_2} & r_{x_2 x_3} \\ r_{x_3 x_1} & r_{x_3 x_2} & r_{x_3 x_3} \end{vmatrix} = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0.$$

Чим ближче до нуля визначник матриці міжфакторної кореляції, тим сильніша мультиколінеарність факторів і ненадійніші результати множинної регресії. І, навпаки, чим ближче до одиниці визначник матриці міжфакторної кореляції, тим менше мультиколінеарність факторів.

Існує низка підходів до подолання сильної міжфакторної кореляції. Найпростіший шлях усунення мультиколінеарності полягає у виключенні з моделі одного або кількох факторів. Інший підхід пов'язаний з перетворенням факторів, за яким зменшується кореляція між ними.

Одним із шляхів урахування внутрішньої кореляції факторів є перехід до сполучених рівнянь регресії, тобто до рівнянь, які відбивають не тільки вплив факторів, але і їхню взаємодію. Так, якщо $\hat{y} = a + b_1 x_1 + b_2 x_2 + b_3 x_m$, то можлива побудова сполученого рівняння наступного вигляду:

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 + \varepsilon.$$

Розглянуте рівняння включає взаємодію першого порядку (взаємодія двох факторів). Можливе включення до моделі і взаємодій більш високого порядку, якщо буде доведена їхня статистична значущість за F-критерієм Фішера, але, як правило, взаємодії третього та більш високих порядків виявляються статистично незначущими.

Відбір факторів, що включаються до регресії, є одним з найважливіших етапів практичного використання методів регресії. Підходи до відбору факторів на основі показників кореляції можуть бути різні. Вони призводять побудову рівняння множинної регресії відповідно до різних методик. Залежно від того, яка методика побудови рівняння регресії прийнята, змінюється алгоритм її рішення на ЕОМ.

При відборі факторів рекомендують також користуватися наступним правилом: число факторів, що включають до моделі, зазвичай у 6-7 разів менше обсягу сукупності, за якою будують регресію. Якщо це співвідношення порушене, то число ступенів свободи залишкової дисперсії дуже мале. Це призводить до того, що параметри рівняння регресії виявляються статистично незначущими, а F-критерій менше табличного значення.

Ще одним з можливих методів усунення або зменшення мультиколінеарності є використання покрокових процедур відбору найбільш інформативних змінних. Наприклад, на першому кроці розглядають лише одну пояснюючу змінну, що має із залежною змінною Y найбільший коефіцієнт детермінації. На другому кроці включають до регресії нову пояснюючу змінну, що разом з відбраною спочатку утворює пару пояснюючих змінних, яка має з Y найбільш високий (скорегований) коефіцієнт детермінації. На третьому кроці вводять до регресії ще одну пояснюючу змінну, котра разом із двома відібраними спочатку утворює трійку пояснюючих змінних, що має з Y найбільший (скорегований) коефіцієнт детермінації і так далі.

Процедура введення нових змінних триває доти, поки буде збільшуватися відповідний (скорегований) коефіцієнт детермінації R^2 . Окрім розглянутої вище покрокової процедури приєднання пояснюючих змінних, використовують також покрокові процедури приєднання-видалення і процедуру видалення пояснюючих змінних. Слід зазначити, що яка би покрокова процедура не використовувалася, вона не гарантує визначення оптимального (у сенсі одержання максимального коефіцієнта детермінації R^2) набору пояснюючих змінних. Однак у більшості випадків одержувані за допомогою покрокових процедур набори змінних виявляються оптимальними або близькими до них.

3.4 Лінійні регресійні моделі з гетероскедастичними залишками

При оцінці параметрів рівняння регресії застосовують метод найменших квадратів (МНК). При цьому відповідно до теореми Гаусса-Маркова виконують певні передумови щодо випадкової складової ε у моделі

$$y = a + b_1x_1 + b_2x_2 + \dots + b_mx_m + \varepsilon$$

випадкова складова ε являє собою неспостережувану величину. Після того як зроблена оцінка параметрів моделі, розраховуючи різниці фактичних і теоретичних значень результативної ознаки y , можна визначити оцінки випадкової складової $y - \hat{y}_x$. Їх можна вважати деякою вибірковою реалізацією невідомого залишку заданого рівняння, тобто ε_i .

При зміні специфікації моделі, додаванні до неї нових спостережень, вибіркові оцінки залишків ε_i можуть змінюватися. Тому в задачу регресійного аналізу входить не тільки побудова самої моделі, але й дослідження випадкових відхилень ε_i , тобто залишкових величин.

При використанні критеріїв Фішера і Стюдента роблять припущення щодо поведінки залишків ε_i , а саме, що вони являють собою незалежні випадкові величини та їхнє математичне сподівання дорівнює 0, що вони мають однакову (постійну) дисперсію і підкоряються нормальному розподілу.

Статистична перевірка параметрів регресії і показників кореляції заснована на неперевірюваних передумовах розподілу випадкової складової ε_i . Вона носить лише попередній характер. Після побудови рівняння регресії проводять перевірку наявності у випадкових залишків ε_i передбачуваних властивостей. Це пов'язане з тим, що оцінки параметрів регресії повинні мати певні властивості. Вони повинні бути незміщеними, спроможними і ефективними.

Незміщеність оцінки означає, що математичне сподівання залишків дорівнює нулю. Якщо оцінки параметрів мають властивість незміщеності, то їх можна порівнювати.

Оцінки вважаються **ефективними**, якщо вони характеризуються найменшою дисперсією. У практичних дослідженнях це означає можливість переходу від точкового оцінювання параметрів до інтервального.

Спроможність оцінок характеризує збільшення їхньої точності зі збільшенням обсягу вибірки. Великий практичний інтерес представляють ті результати регресійного аналізу, для яких довірчий інтервал очікуваного значення параметра регресії b_i має у границі значення імовірності, що дорівнює одиниці. Іншими словами, імовірність одержання оцінки на заданій відстані від істинного значення параметра близька до одиниці.

Метод найменших квадратів визначає оцінки регресії на основі мінімізації суми квадратів залишків. Тому дослідження залишків ε_i припускає перевірку наявності наступних п'яти передумов МНК:

- збурення ε є величина випадкова;
- математичне сподівання збурення ε дорівнює нулю;
- дисперсія збурення ε постійна для будь-якого i - умова гомоскедастичності або рівномірності збурення;
- відсутність автокореляції залишків – значення залишків ε_i розподілені незалежно один від одного;
- збурення ε_i є нормально розподілена випадкова величина.

Якщо розподіл випадкових залишків ε_i не відповідає певним передумовам МНК, то необхідно корегувати модель.

Насамперед, перевіряють випадковий характер залишків ε_i – перша передумова МНК. З цією метою будують графік залежності залишків ε_i від теоретичних значень результативної ознаки (рис. 3.1, а). Якщо графіком є горизонтальна смуга точок, то залишки ε_i являють собою випадкові величини, і теоретичні значення \hat{y}_x добре апроксимують фактичні значення y .

Друга передумова МНК щодо нульового значення математичного сподівання залишків означає, що $\sum (y - \hat{y}_x) = 0$. Це здійснюється для лінійних моделей і моделей, що нелінійні щодо включених змінних.

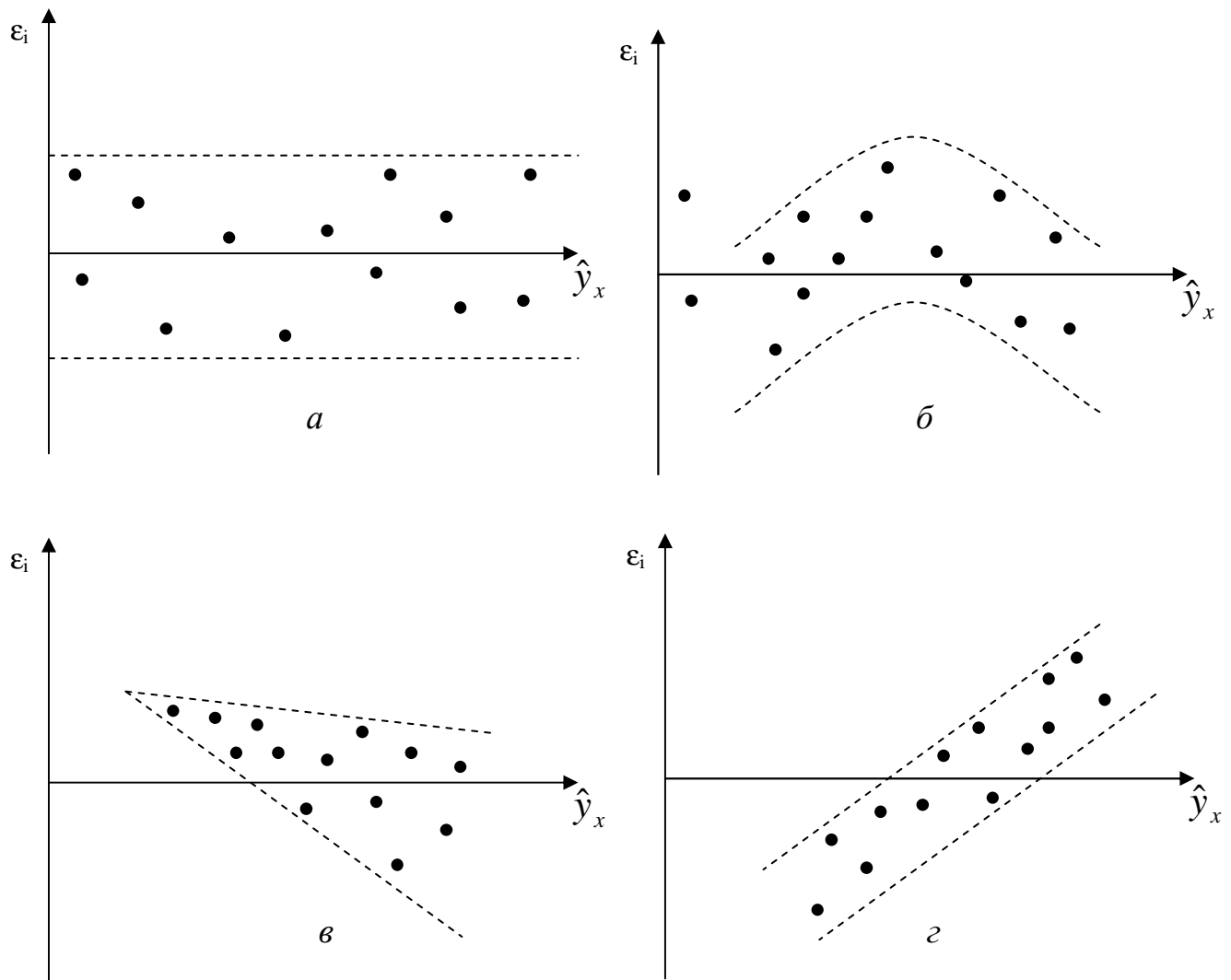


Рисунок 3.1 - Залежність випадкових залишків ε_i від теоретичних значень \hat{y}_x :

а - залишки ε_i є випадковими величинами; б - залишки ε_i не випадкові;

в - залишки ε_i не мають постійної дисперсії;

г - залишки ε_i носять систематичний характер

Разом з тим, незміщеність оцінок коефіцієнтів регресії, отриманих за МНК, залежить від незалежності випадкових залишків і величин x_j , що також досліджують в рамках дотримання другої передумови МНК. З цією метою поряд з розглянутим графіком залежності залишків ε_i від теоретичних значень результативної ознаки \hat{y}_x будують графік залежності випадкових залишків ε_i від факторів, включених до регресії x_j (рис. 3.2).

Якщо залишки на графіку розташовані у вигляді горизонтальної смуги точок, то вони незалежні від значень x_j . Якщо ж графік показує наявність залежності ε_i і x_j , то модель неадекватна. Причини неадекватності можуть бути різ-

ні. Можливо, що порушено третю передумову МНК, і дисперсія залишків не постійна для кожного значення фактору x_j . Може бути невірною специфікація моделі, і до неї необхідно ввести додаткові члени від x_j , наприклад x_j^2 . Скупчення точок у певних ділянках значень фактора x_j говорить про наявність систематичної помилки моделі.

Передумова про нормальний розподіл залишків дозволяє проводити перевірку параметрів регресії за допомогою F- і t -критеріїв. Разом з тим, оцінки параметрів регресії, знайдені із застосуванням МНК, мають гарні властивості навіть при відсутності нормального розподілу залишків, тобто при порушенні п'ятої передумови МНК.

Зовсім необхідним для одержання за МНК спроможних оцінок параметрів регресії є дотримання третьої і четвертої передумов.

Відповідно до третьої передумови МНК потрібно, щоб дисперсія залишків була гомоскедастичною. Це означає, що для кожного значення фактору x_j залишки ε_i мають однакову дисперсію. Якщо ця умова застосування МНК не виконується, то має місце гетероскедастичність. Наявність гетероскедастичності показана на рисунку 3.3, в.

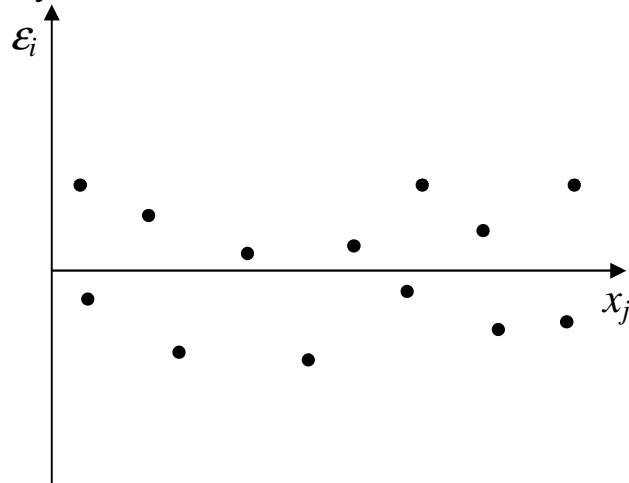


Рисунок 3.2 - Залежність величини залишків ε_i від величини фактору x_j

Для множинної регресії є найбільш прийнятний візуальний спосіб вивчення гомо- і гетероскедастичності на підставі графіків залежності залишків ε_i від теоретичних значень результативної ознаки \hat{y}_x .

При побудові регресійних моделей надзвичайно важливе дотримання четвертої передумови МНК, що полягає у відсутності автокореляції залишків, тобто значення залишків ε_i розподілені незалежно один від одного. Автокореляція залишків означає наявність кореляції між залишками поточних і попередніх (наступних) спостережень. Коефіцієнт кореляції між ε_i і ε_j , де ε_i – залишки поточних спостережень, ε_j – залишки попередніх спостережень (наприклад, $j=i-1$), може бути визначений як

$$r_{\varepsilon_i \varepsilon_j} = \frac{\text{cov}(\varepsilon_i \varepsilon_j)}{\sigma_{\varepsilon_i} \sigma_{\varepsilon_j}},$$

тобто за звичайною формулою лінійного коефіцієнта кореляції. Якщо цей коефіцієнт виявиться істотно відмінним від нуля, то залишки автокорельовані і функція щільності імовірності $f(\varepsilon)$ залежить від j -ї точки спостереження і від розподілу значень залишків в інших точках спостереження.

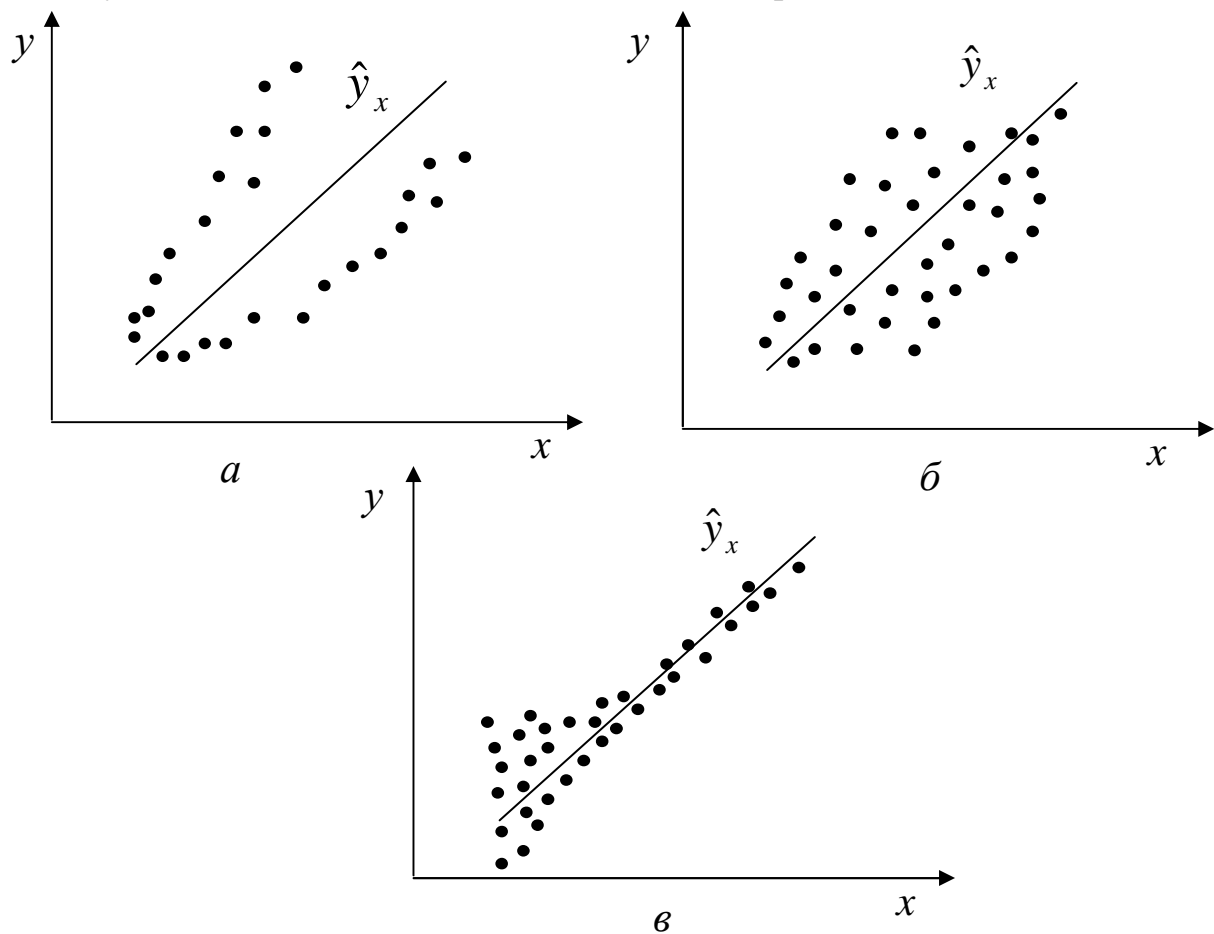


Рисунок 3.3 - Приклади гетероскедастичності:

а - дисперсія залишків зростає в міру збільшення x ; б - дисперсія залишків досягає максимальної величини при середніх значеннях змінної x і зменшується при мінімальних і максимальних значеннях x ; в - максимальна дисперсія залишків при малих значеннях x і стає однорідною в міру збільшення значень x

Відсутність автокореляції залишкових величин забезпечує спроможність і ефективність оцінок коефіцієнтів регресії. Особливо актуальним є дотримання цієї передумови МНК при побудові регресійних моделей за рядами динаміки, де у зв'язку з наявністю тенденції наступні рівні динамічного ряду, як правило, залежать від своїх попередніх рівнів.

При недотриманні основних передумов МНК доводиться корегувати модель, змінюючи її специфікацію, додавати або виключати деякі фактори, перетворювати вихідні дані для того, щоб одержати оцінки коефіцієнтів регресії, які мають властивість незміщеності, мають менше значення дисперсії залишків і забезпечують у зв'язку з цим більш ефективну статистичну перевірку значущості параметрів регресії.

3.5 Узагальнений метод найменших квадратів (УМНК)

При порушенні властивості гомоскедастичності й наявності автокореляції залишків ε_i рекомендують традиційний метод найменших квадратів замінити **узагальненим методом**.

Узагальнений метод найменших квадратів застосовують до перетворених даних, він дозволяє одержувати оцінки, які мають не тільки властивість незміщеності, але й менші вибіркові дисперсії. Зупинимося на використанні УМНК для корегування гетероскедастичності.

Як і раніше, будемо припускати, що середнє значення залишкових величин дорівнює нулю. А от їхня дисперсія не залишається незмінною для різних значень фактора, а пропорційна величині K_i , тобто

$$\sigma_{\varepsilon_i}^2 = \sigma^2 K_i,$$

де $\sigma_{\varepsilon_i}^2$ – дисперсія помилки при конкретному i -му значенні фактору;

σ^2 – постійна дисперсія помилки при дотриманні передумови про гомоскедастичність залишків;

K_i – коефіцієнт пропорційності, що змінюється зі зміною величини фактору, що й зумовлює неоднорідність дисперсії.

При цьому вважають, що σ^2 невідома, а відносно величин K_i висувують певні гіпотези, які характеризують структуру гетероскедастичності.

У загальному виді для рівняння $y_i = a + bx_i + \varepsilon_i$ при $\sigma_{\varepsilon_i}^2 = \sigma^2 K_i$ модель прийме вигляд: $y_i = a + bx_i + \sqrt{K_i} \varepsilon_i$. В ній залишкові величини є гетероскедастичними. Припускаючи в них відсутність автокореляції, можна перейти до рівняння з гомоскедастичними залишками, розділивши всі змінні, зафіксовані в ході i -го спостереження, на $\sqrt{K_i}$. Тоді дисперсія залишків буде величиною постійною, тобто $\sigma_{\varepsilon_i}^2 = \sigma^2$.

Іншими словами, від регресії y на x ми перейдемо до регресії на нові змінні: $\frac{y}{\sqrt{K}}$ і $\frac{x}{\sqrt{K}}$. Рівняння регресії прийме вигляд:

$$\frac{y_i}{\sqrt{K_i}} = \frac{a}{\sqrt{K_i}} + b \frac{x_i}{\sqrt{K_i}} + \varepsilon_i,$$

а вихідні дані для даного рівняння матимуть вигляд:

$$y = \begin{pmatrix} \frac{y_1}{\sqrt{K_1}} \\ \frac{y_2}{\sqrt{K_2}} \\ \dots \\ \frac{y_n}{\sqrt{K_n}} \end{pmatrix}, \quad x = \begin{pmatrix} \frac{x_1}{\sqrt{K_1}} \\ \frac{x_2}{\sqrt{K_2}} \\ \dots \\ \frac{x_n}{\sqrt{K_n}} \end{pmatrix}.$$

Стосовно звичайної регресії рівняння з новими перетвореними змінними являє собою зважену регресію, в якій змінні y і x взяті з вагами $\frac{1}{\sqrt{K}}$.

Оцінка параметрів нового рівняння з перетвореними змінними призводить до зваженого методу найменших квадратів, для якого необхідно мінімізувати суму квадратів відхилень виду

$$S(a, b) = \sum_{i=1}^n \frac{1}{K_i} (y_i - a - bx_i)^2.$$

Відповідно дістанемо наступну систему нормальних рівнянь:

$$\begin{cases} \sum_{i=1}^n \frac{y_i}{K_i} = a \sum_{i=1}^n \frac{1}{K_i} + b \sum_{i=1}^n \frac{x_i}{K_i} \\ \sum_{i=1}^n \frac{y_i x_i}{K_i} = a \sum_{i=1}^n \frac{x_i}{K_i} + b \sum_{i=1}^n \frac{x_i^2}{K_i} \end{cases}.$$

Якщо перетворені змінні x і y взяті у відхиленнях від середніх рівнів, то коефіцієнт регресії b можна визначити як

$$b = \frac{\sum_{i=1}^n \frac{1}{K_i} x_i y_i}{\sum_{i=1}^n \frac{1}{K_i} x_i^2}.$$

При звичайному застосуванні методу найменших квадратів до рівняння лінійної регресії для змінних у відхиленнях від середніх рівнів коефіцієнт регресії b визначають за формулою:

$$b = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Як бачимо, при використанні узагальненого МНК з метою корегування гетероскедастичності коефіцієнт регресії b являє собою зважену величину щодо звичайного МНК з вагою $\frac{1}{\sqrt{K}}$.

Аналогічний підхід можливий не тільки для рівняння парної, але і для множинної регресії. Припустимо, що розглядають модель виду

$$y = a + b_1 x_1 + b_2 x_2 + \varepsilon,$$

для якої дисперсія залишкових величин виявилася пропорційною K_i^2 . K_i являє собою коефіцієнт пропорційності, що приймає різні значення для відповідних i значень факторів x_1 і x_2 . Через те, що $\sigma_{\varepsilon_i}^2 = \sigma^2 K_i^2$, розглянута модель прийме вид

$$y_i = a + b_1 x_{1i} + b_2 x_{2i} + K_i \varepsilon_i,$$

де залишки ε_i є гетероскедастичними.

Для того щоб одержати рівняння, де залишки ε_i є гомоскедастичними, слід перейти до нових перетворених змінних, розділивши всі члени вихідного рівняння на коефіцієнт пропорційності K . Рівняння з перетвореними змінними матиме вигляд

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i.$$

Це рівняння не містить вільного члена. Разом з тим, знайшовши змінні в новому перетвореному вигляді і застосовуючи звичайний МНК до них, одержимо іншу специфікацію моделі:

$$\frac{y_i}{K_i} = A + b_1 \frac{x_{1i}}{K_i} + b_2 \frac{x_{2i}}{K_i} + \varepsilon_i.$$

Параметри такої моделі залежать від концепції, прийнятої для коефіцієнта пропорційності K_i . В економетричних дослідженнях досить часто пропонують гіпотезу, що залишки ε_i пропорційні значенням фактору. Так, якщо у рівнянні

$$y = a + b_1 x_1 + b_2 x_2 + \dots + b_m x_m + e$$

припустити, що $e = \varepsilon x_1$, тобто $K = x_1$ і $\sigma_{\varepsilon_i}^2 = \sigma^2 x_1$, то узагальнений МНК припускає оцінку параметрів наступного трансформованого рівняння:

$$\frac{y}{x_1} = b_1 + b_2 \frac{x_2}{x_1} + \dots + b_m \frac{x_m}{x_1} + \varepsilon.$$

Застосування в цьому випадку узагальненого МНК призводить до того, що спостереження з меншими значеннями перетворених змінних $\frac{x}{K}$ мають при визначенні параметрів регресії відносно більшу вагу, ніж з первісними змінними. Разом з тим, варто мати на увазі, що нові перетворені змінні одержують но-

вий економічний зміст, і їхня регресія має інший зміст, ніж регресія за вихідними даними.

Розглянемо приклад. Нехай y – витрати виробництва, x_1 – обсяг продукції, x_2 – основні виробничі фонди, x_3 – чисельність працівників. Тоді рівняння

$$y = a + b_1 x_1 + b_2 x_2 + b_3 x_3 + e$$

є моделлю витрат виробництва з об'ємними факторами. Припускаючи, що $\sigma_{\varepsilon_i}^2$ пропорційна квадрату чисельності працівників x_3 , ми одержимо як результативну ознаку витрати на одного працівника $\frac{y}{x_3}$, а як фактори - наступні показники: продуктивність праці $\frac{x_1}{x_3}$ і фондоозброєність праці $\frac{x_2}{x_3}$. Відповідно трансформована модель прийме вигляд

$$\frac{y}{x_3} = b_3 + b_1 \frac{x_1}{x_3} + b_2 \frac{x_2}{x_3} + \varepsilon,$$

де параметри b_1 , b_2 , b_3 чисельно не збігаються з аналогічними параметрами попередньої моделі. Окрім цього, коефіцієнти регресії змінюють економічний зміст: з показників сили зв'язку, що характеризують середню абсолютну зміну витрат виробництва зі зміною абсолютної величини відповідного фактора на одиницю, вони фіксують при узагальненому МНК середню зміну витрат на працівника; зі зміною продуктивності праці на одиницю при незмінному рівні фондоозброєності праці; і зі зміною фондоозброєності праці на одиницю при незмінному рівні продуктивності праці.

Якщо припустити, що в моделі з первісними змінними дисперсія залишків пропорційна квадрату обсягу продукції, $\sigma_{\varepsilon_i}^2 = \sigma^2 x_1^2$, можна перейти до рівняння регресії виду

$$\frac{y}{x_1} = b_1 + b_2 \frac{x_2}{x_1} + b_3 \frac{x_3}{x_1} + \varepsilon.$$

У ньому нові змінні: $\frac{y}{x_1}$ – витрати на одиницю (або на 1 грн. продукції),

$\frac{x_2}{x_1}$ – фондоємність продукції, $\frac{x_3}{x_1}$ – трудоємність продукції.

Гіпотеза про те, що залишки пропорційні величині фактору, може мати реальну підставу: при обробці недостатньо однорідної сукупності, що включає як великі, так і малі підприємства, більшим об'ємним значенням фактора можуть відповідати більші дисперсія результативної ознаки і дисперсія залишкових величин.

За наявності однієї пояснюючої змінної гіпотеза $\sigma_{\varepsilon_i}^2 = \sigma^2 x^2$ трансформує лінійне рівняння

$$y = a + bx + e$$

на рівняння

$$\frac{y}{x} = b + \frac{a}{x} + \varepsilon,$$

в якому параметри a і b помінялися місцями, константа стала коефіцієнтом нахилу лінії регресії, а коефіцієнт регресії - вільним членом.

Перехід до відносних величин істотно знижує варіацію фактору і відповідно зменшує дисперсію помилки. Він являє собою найбільш простий випадок урахування гетероскедастичності в регресійних моделях за допомогою узагальненого МНК. Процес переходу до відносних величин може бути ускладнений висуванням інших гіпотез про пропорційність помилок щодо включених у модель факторів. Використання тієї або іншої гіпотези припускає спеціальні дослідження залишкових величин для відповідних регресійних моделей. Застосування узагальненого МНК дозволяє одержати оцінки параметрів моделі, що мають меншу дисперсію.

3.6 Регресійні моделі із змінною структурою. Фіктивні змінні

Дотепер як фактори розглядалися економічні змінні, що приймають кількісні значення в певному інтервалі. Разом з тим, може виявитися необхідним включити до моделі фактор, що має два або більше якісних рівнів. Це можуть бути різного роду атрибутивні ознаки, такі, наприклад, як професія, стать, освіта, кліматичні умови, сезонність, належність до певного регіону. Якісні ознаки можуть істотно впливати на структуру лінійних зв'язків між змінними і призводити до стрибкоподібної зміни параметрів регресійної моделі. В цьому випадку говорять про дослідження регресійних моделей зі змінною структурою.

Щоб ввести такі змінні в регресійну модель, їм треба привласнити ті або інші цифрові мітки, тобто якісні змінні перетворюють на кількісні. Такого виду сконструйовані змінні в економетриці прийнято називати **фіктивними змінними**.

Розглянемо застосування фіктивних змінних для функції попиту. Припустимо, що за групою осіб чоловічої й жіночої статі вивчається лінійна залежність споживання кави від ціни. У загальному виді для сукупності обстежуваних рівняння регресії має вигляд:

$$y = a + bx + \varepsilon,$$

де y - кількість споживаної кави;

x - ціна.

Аналогічні рівняння можуть бути знайдені окремо для осіб чоловічої статі: $y_1 = a_1 + b_1 x_1 + \varepsilon_1$ і жіночої статі: $y_2 = a_2 + b_2 x_2 + \varepsilon_2$.

Розходження у споживанні кави виявляться у розходженні середніх \bar{y}_1 і \bar{y}_2 . Разом з тим, сила впливу x на y може бути однаковою, тобто $b \approx b_1 \approx b_2$. В

цьому випадку можлива побудова загального рівняння регресії з включенням до нього фактора «стать» у вигляді фіктивної змінної. Поєднуючи рівняння y_1 і y_2 і вводячи фіктивні змінні, можна прийти до наступного виразу:

$$y = a_1 z_1 + a_2 z_2 + bx + \varepsilon,$$

де z_1 і z_2 – фіктивні змінні, що приймають значення:

$$z_1 = \begin{cases} 1 - \text{чоловіча стаття}, \\ 0 - \text{жіноча стаття} \end{cases}; \quad z_2 = \begin{cases} 0 - \text{чоловіча стаття}, \\ 1 - \text{жіноча стаття} \end{cases}.$$

У загальному рівнянні регресії залежна змінна y розглядається як функція не тільки ціни x , але і статі (z_1, z_2). Змінна z розглядається як дихотомічна змінна, яка приймає всього два значення: 1 і 0. При цьому, коли $z_1=1$, то $z_2=0$, і навпаки.

Для осіб чоловічої статі, коли $z_1=1$ і $z_2=0$, об'єднане рівняння регресії матиме вигляд: $\hat{y} = a_1 + bx$, а для осіб жіночої статі, коли $z_1=0$ і $z_2=1$: $\hat{y} = a_2 + bx$. Іншими словами, розходження в споживанні для осіб чоловічої і жіночої статі викликані розходженнями вільних членів рівняння регресії: $a_1 \neq a_2$. Параметр b є загальним для всієї сукупності осіб, як для чоловіків, так і для жінок.

Проте при введенні двох фіктивних змінних z_1 і z_2 у модель $y = a_1 z_1 + a_2 z_2 + bx + \varepsilon$ застосування МНК для оцінювання параметрів a_1 і a_2 призведе до виродженої матриці вихідних даних, а отже, і до неможливості одержання їхніх оцінок. Пояснюють це тим, що при використанні МНК у даному рівнянні з'являється вільний член, тобто рівняння прийме вигляд

$$y = A + a_1 z_1 + a_2 z_2 + bx + \varepsilon.$$

Припускаючи при параметрі A незалежну змінну рівною 1, маємо матрицю вихідних даних:

$$\begin{bmatrix} 1 & 1 & 0 & x_1 \\ 1 & 1 & 0 & x_2 \\ 1 & 0 & 1 & x_3 \\ 1 & 1 & 0 & x_4 \\ \dots & \dots & \dots & \dots \\ 1 & 0 & 1 & x_n \end{bmatrix}$$

У розглянутій матриці є лінійна залежність між першим, другим і третім стовпцями: перший дорівнює сумі другого і третього стовпців. Тому матриця вихідних факторів є виродженою. Виходом з утруднення, що створилося, може з'явитися перехід до рівнянь

$$y = A + A_1 z_1 + bx + \varepsilon \quad \text{або} \quad y = A + A_2 z_2 + bx + \varepsilon,$$

тобто кожне рівняння включає тільки одну фіктивну змінну z_1 або z_2 .

Припустимо, що визначено рівняння

$$y = A + A_1 z_1 + bx + \varepsilon,$$

де z_1 приймає значення 1 для чоловіків і 0 для жінок.

Теоретичні значення об'єму споживання кави для чоловіків будуть отримані з рівняння

$$\hat{y} = A + A_1 + bx.$$

Для жінок відповідні значення одержимо з рівняння

$$\hat{y} = A + bx.$$

Зіставляючи ці результати, бачимо, що розходження у рівні споживання чоловіками й жінками полягають у розходженні вільних членів даних рівнянь: A – для жінок і $A + A_1$ – для чоловіків.

Тепер якісний фактор приймає тільки два стани, яким відповідають значення 1 і 0. Якщо ж число градацій якісної ознаки-фактору перевищує два, то до моделі вводять кілька фіктивних змінних, число яких має бути меншим за число якісних градацій. Тільки при дотриманні цього положення матриця вихідних фіктивних змінних не матиме лінійно залежних стовпців, і можлива оцінка параметрів моделі.

В окремих випадках може виявитися необхідним введення двох і більше груп фіктивних змінних, тобто двох і більше якісних факторів, кожний з яких може мати кілька градацій. Наприклад, при вивченні споживання певного товару поряд з факторами, що мають кількісне вираження (ціна, дохід на одного члена родини, ціна на взаємозамінні товари та ін.), враховують і якісні фактори. За їхньою допомогою оцінюють розходження в споживанні окремими соціальними групами населення, диференціацію в споживанні за статтю, національним складом та ін. При побудові такої моделі з кожної групи фіктивних змінних слід виключити по одній змінній. Так, якщо модель включатиме три соціальні групи, три вікові категорії і ряд економічних змінних, то вона прийме вид:

$$y = a + b_1 s_1 + b_2 s_2 + b_3 z_1 + b_4 z_2 + b_5 x_1 + b_6 x_2 + \dots + b_{m+4} x_m + \varepsilon,$$

де y - споживання;

$$s_i = \begin{cases} 1 - \text{якщо спостереження належать до } i - \text{ї соціальної групи,} \\ 0 - \text{в інших випадках,} \end{cases}$$

$$z_i = \begin{cases} 0 - \text{якщо спостереження належать до } i - \text{ї групи за віком,} \\ 1 - \text{в інших випадках,} \end{cases}$$

x_1, x_2, x_m – економічні (кількісні) змінні.

Досі ми розглядали фіктивні змінні як фактори, які використовують в регресійній моделі поряд з кількісними змінними. Разом з тим, можлива регресія тільки на фіктивні змінні. Наприклад, вивчають диференціацію заробітної плати робітників високої кваліфікації за регіонами країни. Модель заробітної плати може мати вигляд:

$$\hat{y} = a + b_1 z_1 + b_2 z_2 + \dots + b_m z_m,$$

де y - середня заробітна плата робітників високої кваліфікації за окремими підприємствами;

$$z_1 = \begin{cases} 1 - \text{якщо підприємство розташовано в східному регіоні,} \\ 0 - \text{в інших регіонах,} \end{cases}$$

$$z_2 = \begin{cases} 1 - \text{якщо підприємство розташовано в західному регіоні,} \\ 0 - \text{в інших регіонах,} \end{cases}$$

.....

$$z_m = \begin{cases} 1 - \text{якщо підприємство розташовано в центральному регіоні,} \\ 0 - \text{в інших регіонах.} \end{cases}$$

Ми розглянули моделі з фіктивними змінними, в яких останні виступають факторами. Може виникнути необхідність побудувати модель, в якій дихотомічна ознака, тобто ознака, що може приймати тільки два значення, відіграє роль результативної. Подібного виду моделі застосовують, наприклад, при обробці даних соціологічних опитувань. Як залежну змінну у розглядають відповіді на запитання, надані в альтернативній формі: «так» або «ні». Тому залежна змінна має два значення: 1 - коли має місце відповідь «так», і 0 - у всіх інших випадках. Модель такої залежної змінної має вигляд:

$$y = a + b_1x_1 + \dots + b_mx_m + \varepsilon.$$

Модель є імовірнісною лінійною моделлю. В ній у приймає значення 1 і 0, яким відповідають імовірності p і $1-p$. Тому при визначенні параметрів моделі знаходять оцінку умовної імовірності події у при фіксованих значеннях x . Для оцінки параметрів лінійно-імовірнісної моделі застосовують методи Logit-, Probit- і Tobit-аналізу. Такого роду моделі використовують при роботі з некілкісними змінними. Як правило, це моделі вибору із заданого набору альтернатив. Залежна змінна у представлена дискретними значеннями (набір альтернатив), що пояснюють змінні x_j – характеристики альтернатив (час, ціна), z_j – характеристики індивідів (вік, дохід, рівень освіти). Модель такого роду дозволяє прогнозувати частку індивідів у генеральній сукупності, які вибирають дану альтернативу.

Серед моделей з фіктивними змінними найбільші прогностичні можливості мають моделі, в яких залежна змінна у розглядається як функція ряду економічних факторів x_i і фіктивних змінних z_j . Останні звичайно відбивають розходження у формуванні результативної ознаки за окремими групами одиниць сукупності, тобто в результаті неоднорідної структури просторового або часового характеру.

Запитання для самоперевірки

1. У чому полягає специфікація моделі множинної регресії?
2. Сформулюйте вимоги до факторів щодо включення їх до моделі множинної регресії.
3. До яких ускладнень призводить мультиколінеарність факторів? Як її подолати?

4. Які коефіцієнти використовують для оцінювання порівняльного впливу факторів на результативну ознаку?
5. Яке призначення часткової кореляції під час побудови моделі множинної кореляції?
6. Як пов'язані один з одним t -критерій Стюдента для оцінювання значущості b_i та часткові F -критерії?
7. Охарактеризуйте основні посилення щодо застосування методу найменших квадратів для побудови регресійної моделі.
8. Охарактеризуйте сутність аналізу залишків регресійної моделі.
9. Як перевірити наявність гомо- або гетероскедастичності залишків?
10. Сформулюйте умови застосування узагальненого методу найменших квадратів.
11. Якими властивості повинні мати оцінки параметрів регресії?
12. Як впливає на параметри множинної лінійної моделі порушення умови, що математичне сподівання збурювання ϵ дорівнює нулю?
13. Як впливає на параметри множинної лінійної моделі порушення умови, що дисперсія збурювання ϵ постійною?
14. Який спосіб використовують для вивчення гомо- і гетероскедастичності?
15. На підставі якого дослідження роблять висновок про автокорельованість залишків?
16. Як діють при недотриманні основних передумов МНК?
17. Чим відрізняється узагальнений МНК від звичайного? У яких випадках його використовують?
18. Які властивості притаманні оцінкам параметрів моделі, отриманим на основі узагальненого МНК?

ЗМ 2 ЕКОНОМЕТРИЧНІ МОДЕЛІ ДИНАМІКИ. СИСТЕМА СТРУКТУРНИХ РІВНЯНЬ

Тема 4 ЧАСОВІ РЯДИ І ПРОГНОЗУВАННЯ

4.1 Загальні відомості про часові ряди і завдання їхнього аналізу

При розгляді класичної моделі регресії характер експериментальних даних, як правило, не має принципового значення. Однак це виявляється не так, коли порушені умови класичної моделі. Методи дослідження моделей, заснованих на даних просторових вибірок і часових рядів істотно відрізняються. Пояснюється це тим, що на відміну від просторових вибірок спостереження в часових рядах, як правило, не можна вважати незалежними.

Часовий ряд (ряд динаміки) – це сукупність значень будь-якого показника за кілька послідовних моментів або періодів часу. Кожний рівень часового ряду формується під впливом великої кількості факторів, які умовно можна підрозділити на три групи:

- фактори, що формують тенденцію ряду;
- фактори, що формують циклічні коливання ряду;
- випадкові фактори.

Більшість часових рядів економічних показників мають тенденцію, що характеризує сукупний довгостроковий вплив множини факторів на динаміку досліджуваного показника. Всі ці фактори, взяті окремо, можуть робити різнонаправлений вплив на досліджуваний показник. Однак у сукупності вони формують його зростаючу або спадну тенденцію. На рисунку 4.1 показаний гіпотетичний часовий ряд, що містить зростаючу тенденцію.

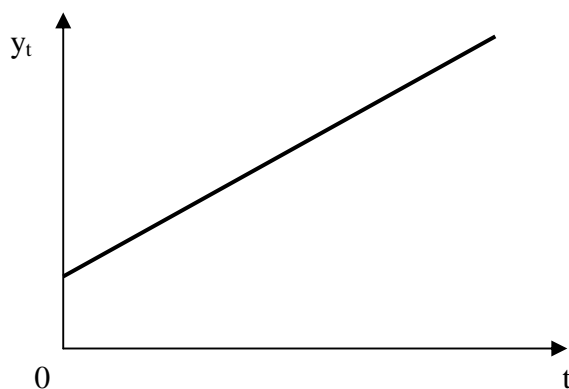


Рисунок 4.1 - Часовий ряд, що містить зростаючу тенденцію

Досліджуваний показник може також бути підданий циклічним коливанням. Ці коливання можуть носити сезонний характер, оскільки економічна діяльність ряду галузей економіки залежить від пори року (наприклад, ціни на сільськогосподарську продукцію в літній період нижчі, ніж у зимовий; рівень безробіття в курортних містах у зимовий період вищий порівняно з літнім). При

наявності великих масивів даних за тривалі проміжки часу можна виявити циклічні коливання, пов'язані із загальною динамікою кон'юнктури ринку. На рисунку 4.2 поданий гіпотетичний часовий ряд, що містить тільки сезонну компоненту.

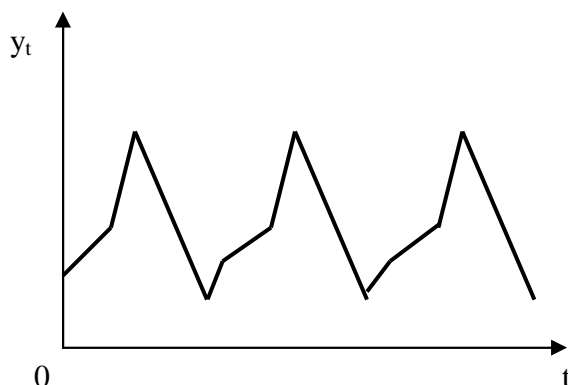


Рисунок 4.2 - Часовий ряд, що містить тільки сезонну компоненту

Деякі часові ряди не містять тенденції і циклічної компоненти, а кожний наступний їхній рівень утворюється як сума середнього рівня ряду і якоїсь (додатної або від'ємної) випадкової компоненти. Приклад ряду, що містить тільки випадкову компоненту, наведений на рисунку 4.3.

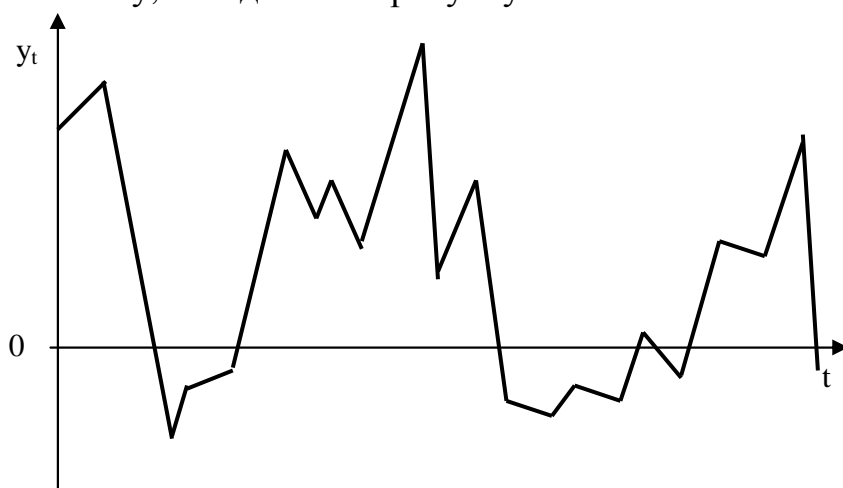


Рисунок 4.3 - Часовий ряд, що містить тільки випадкову компоненту

Очевидно, що реальні дані не впливають цілком і повністю з будь-яких описаних вище моделей. Найчастіше вони містять всі три компоненти. Кожний їхній рівень формується під впливом тенденції, сезонних коливань і випадкової компоненти.

У більшості випадків фактичний рівень часового ряду можна уявити як суму або добуток трендової, циклічної і випадкової компонентів. Модель, в якій часовий ряд представлений як сума перелічених компонентів, називається **адитивною моделлю** часового ряду. Модель, в якій часовий ряд представлений як добуток перелічених компонентів, називається **мультиплікативною моделлю** часового ряду. Основна задача економетричного дослідження окремого часового ряду - виявлення і додання кількісного виразу кожної з перелічених ви-

ще компонентів для того, щоб використати отриману інформацію для прогнозування майбутніх значень ряду або при побудові моделей взаємозв'язку двох або більше часових рядів.

Найважливішою класичною задачею при дослідженні економічних часових рядів є виявлення і статистична оцінка основної тенденції розвитку досліджуваного процесу і відхилень від неї.

Відзначимо основні етапи аналізу часових рядів:

- графічне подання і опис поведінки часового ряду;
- виділення й видалення закономірних (невипадкових) складових часового ряду (тренду, сезонних і циклічних складових);
- згладжування і фільтрація (видалення низько- або високочастотних складових часового ряду);
- дослідження випадкової складової часового ряду, побудова й перевірка адекватності математичної моделі для його опису;
- прогнозування розвитку досліджуваного процесу на основі наявного часового ряду;
- дослідження взаємозв'язку між різними часовими рядами.

Серед найпоширеніших методів аналізу часових рядів виділяють кореляційний і спектральний аналіз, моделі авторегресії і ковзної середньої.

Якщо вибірку y_1, y_2, \dots, y_n розглядають як одну з реалізацій випадкової величини Y , часовий ряд y_1, y_2, \dots, y_n розглядається як одна з реалізацій (траєкторій) випадкового процесу $Y(t)$. Разом з тим необхідно мати на увазі принципи відмінності часового ряду y_t , ($t=1, n$) від послідовності спостережень y_1, y_2, \dots, y_n , що утворюють випадкову вибірку. По-перше, на відміну від елементів випадкової вибірки члени часового ряду, як правило, не є статистично незалежними. По-друге, члени часового ряду не є однаково розподіленими.

4.2 Автокореляція рівнів часового ряду

При наявності в часовому ряді тенденції і циклічних коливань значення кожного наступного рівня ряду залежать від попередніх. Кореляційну залежність між послідовними рівнями часового ряду називають автокореляцією рівнів ряду.

Кількісно її можна виміряти за допомогою лінійного коефіцієнта кореляції між рівнями вихідного часового ряду й рівнями цього ряду, зрушеними на кілька кроків у часі.

Формула для розрахунку коефіцієнта автокореляції має вигляд:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (4.1)$$

$$\text{де } \bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^n y_t, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^n y_{t-1}$$

Цю величину називають коефіцієнтом автокореляції рівнів ряду першого порядку, тому що він вимірює залежність між сусідніми рівнями ряду y_t і y_{t-1} .

Аналогічно можна визначити коефіцієнти автокореляції другого і більш високих порядків. Так, коефіцієнт автокореляції другого порядку характеризує тісноту зв'язку між рівнями y_t і y_{t-2} , і визначають його за формулою:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}}, \quad (4.2)$$

$$\text{де } \bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^n y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^n y_{t-2}.$$

Число періодів, за якими розраховується коефіцієнт автокореляції, називають **лагом**. Із збільшенням лагу число пар значень, за якими розраховують коефіцієнт автокореляції, зменшується. Вважають за доцільне для забезпечення статистичної вірогідності коефіцієнтів автокореляції використовувати правило – макси-

мальний лаг має не перевищувати $\tau \leq \frac{n}{4}$.

Властивості коефіцієнта автокореляції.

1. Його будують за аналогією з лінійним коефіцієнтом кореляції і в такий спосіб він характеризує тісноту тільки лінійного зв'язку поточного і попереднього рівнів ряду. Тому з коефіцієнта автокореляції можна судити про наявність лінійної (або близької до лінійної) тенденції. Для деяких часових рядів, що мають сильну нелінійну тенденцію (наприклад, параболу другого порядку або експоненту), коефіцієнт автокореляції рівнів вихідного ряду може наближатися до нуля.

2. За знаком коефіцієнта автокореляції не можна робити висновок про зростаючу або спадну тенденцію в рівнях ряду. Більшість часових рядів економічних даних містять додатну автокореляцію рівнів, однак при цьому можуть мати спадну тенденцію.

Послідовність коефіцієнтів автокореляції рівнів першого, другого і т. д. порядків називають **автокореляційною функцією** часового ряду. Графік залежності її значень від величини лагу (порядку коефіцієнта автокореляції) називають **корелограмою**.

Аналіз автокореляційної функції і корелограми дозволяє визначити лаг, при якому автокореляція найбільш висока, а отже, і лаг, при якому зв'язок між поточним і попереднім рівнями ряду найбільш тісний, тобто за допомогою аналізу автокореляційної функції і корелограми можна виявити структуру ряду.

Якщо найбільш високим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить тільки тенденцію. Якщо найбільш високим виявився коефіцієнт автокореляції порядку τ , то ряд містить циклічні коли-

вання з періодичністю в τ моментів часу. Якщо жоден з коефіцієнтів автокореляції не є значущим, можна зробити одне з двох припущень щодо структури цього ряду: або ряд не містить тенденції і циклічних коливань, або ряд містить сильну нелінійну тенденцію, для виявлення якої потрібно провести додатковий аналіз. Тому коефіцієнт автокореляції рівнів і автокореляційну функцію доцільно використовувати для виявлення в часовому ряді наявності або відсутності трендової компоненти і циклічної (сезонної) компоненти.

4.3 Моделювання тенденції часового ряду

Розповсюдженим способом моделювання тенденції часового ряду є побудова аналітичної функції, що характеризує залежність рівнів ряду від часу, або тренду. Цей спосіб називають **аналітичним вирівнюванням часового ряду**.

Оскільки залежність від часу може приймати різні форми, для її формалізації можна використати різні види функцій. Для побудови трендів найчастіше застосовують наступні функції:

лінійний тренд: $\hat{y}_t = a + bt$;

гіперболічний: $\hat{y}_t = a + \frac{b}{t}$;

експонентний тренд: $\hat{y}_t = e^{a+bt}$ (або $\hat{y}_t = ab^t$);

степенева функція: $\hat{y}_t = a + t^b$;

поліноми різних ступенів: $\hat{y}_t = a + b_1t + b_2t^2 + \dots + b_mt^m$.

Параметри кожного з перерахованих вище трендів можна визначити звичайним методом найменших квадратів (МНК), використовуючи як незалежну змінну час $t=1, 2, \dots, n$, а як залежну змінну – фактичні рівні часового ряду \hat{y}_t .

$$y_t = f(t) + \varepsilon_t, \quad (4.3)$$

де ε_t – збурення, що задовольняють основним передумовам регресійного аналізу, тобто уявляють собою незалежні і однаково розподілені випадкові величини, розподіл яких передбачається нормальним.

Нагадаємо, що відповідно до МНК параметри лінійного тренду $\hat{y}_t = f(t) = b_0 + b_1t$ знаходять із системи нормальних рівнянь, в якій як пояснююча змінна x_i фігурує час t .

$$\begin{cases} b_0n + b_1 \sum_{t=1}^n t = \sum_{i=1}^n y_i, \\ b_0 \sum_{t=1}^n t + b_1 \sum_{t=1}^n t^2 = \sum_{i=1}^n y_i t. \end{cases} \quad (4.4)$$

При застосуванні МНК для оцінки параметрів експонентної або логістичної функцій виникають складності з розв'язанням системи нормальних рівнянь, тому попередньо, до одержання відповідної системи, звертаються до перетво-

рення цих функцій. Для нелінійних трендів попередньо проводять стандартну процедуру їхньої лінеаризації.

Іншим методом вирівнювання (згладжування) часового ряду, тобто виділення не випадкової складової, є метод ковзних середніх. Він заснований на переході від початкових значень членів ряду до їхніх середніх значень на інтервалі часу, довжина якого визначена заздалегідь. При цьому сам обраний інтервал часу «ковзає» уздовж ряду.

Одержаний у такий спосіб ряд ковзних середніх є більш гладким, ніж вихідний ряд, через усереднення відхилень ряду. Дійсно, якщо індивідуальний розкид значень члену часового ряду y_i , біля свого середнього (згладженого) значення a характеризується дисперсією σ^2 , то розкид середньої з m членів часового ряду $(y_1 + y_2 + \dots + y_m)/m$ біля того самого значення a буде характеризуватися істотно меншою величиною дисперсії, що дорівнює $\frac{\sigma^2}{m}$. Для усереднення може бути використана середня арифметична (проста й з певними вагами), медіана та ін.

Існує кілька способів визначення типу тенденції. До числа найпоширеніших способів належать якісний аналіз досліджуваного процесу, побудова й візуальний аналіз графіку залежності рівнів ряду від часу. З цією самою метою можна використати й коефіцієнти автокореляції рівнів ряду. Тип тенденції можна визначити шляхом порівняння коефіцієнтів автокореляції першого порядку, розрахованих за вихідними і перетвореними рівнями ряду. Якщо часовий ряд має лінійну тенденцію, то його сусідні рівні \hat{y}_t і \hat{y}_{t-1} тісно корелюють. В цьому випадку коефіцієнт автокореляції першого порядку рівнів вихідного ряду має бути високим. Якщо часовий ряд містить нелінійну тенденцію, наприклад, у формі експоненти, то коефіцієнт автокореляції першого порядку за логарифмами рівнів вихідного ряду буде вище за відповідний коефіцієнт, розрахований за рівнями ряду. Чим чіткіше виражена нелінійна тенденція в досліджуваному часовому ряду, тим більшою мірою будуть розрізнятися значення зазначених коефіцієнтів.

Вибір найкращого рівняння у випадку, коли ряд містить нелінійну тенденцію, можна здійснити шляхом перебору основних форм тренду, розрахунку за кожним рівнянням скорегованого коефіцієнта детермінації і середньої помилки апроксимації. Цей метод легко реалізують при комп'ютерній обробці даних.

4.4 Моделювання сезонних коливань

Найпростіший підхід до моделювання сезонних коливань - це розрахунок значень сезонної компоненти методом ковзної середньої і побудова адитивної або мультиплікативної моделі часового ряду.

Загальний вигляд адитивної моделі наступний:

$$Y = T + S + E. \quad (4.5)$$

Ця модель припускає, що кожний рівень часового ряду може бути представлений як сума трендової (T), сезонної (S) і випадкової (E) компонент.

Загальний вид мультиплікативної моделі виглядає в такий спосіб:

$$Y = T * S * E. \quad (4.6)$$

Ця модель припускає, що кожний рівень часового ряду може бути представлений як добуток трендової (T), сезонної (S) і випадкової (E) компонент.

Вибір однієї з двох моделей здійснюють на основі аналізу структури сезонних коливань. Якщо амплітуда коливань приблизно постійна, будують адитивну модель часового ряду, в якій значення сезонної компоненти передбачають постійними для різних циклів. Якщо амплітуда сезонних коливань зростає або зменшується, будують мультиплікативну модель часового ряду, що ставить рівні ряду в залежність від значень сезонної компоненти.

Побудову адитивної і мультиплікативної моделей зводять до розрахунку значень T , S і E для кожного рівня ряду.

Процес побудови моделі містить в собі наступні етапи:

- вирівнювання вихідного ряду методом ковзної середньої;
- розрахунок значень сезонної компоненти S ;
- усунення сезонної компоненти з вихідних рівнів ряду і одержання вирівняних даних ($T+E$) в адитивній або ($T * E$) в мультиплікативній моделі;
- аналітичне вирівнювання рівнів ($T+E$) або ($T * E$) і розрахунок значень T з використанням отриманого рівняння тренду;
- розрахунок отриманих за моделлю значень ($T+E$) або ($T * E$);
- розрахунок абсолютних або відносних помилок. Якщо отримані значення помилок не містять автокореляції, ними можна замінити вихідні рівні ряду і надалі використати часовий ряд помилок E для аналізу взаємозв'язку вихідного ряду та інших часових рядів.

В моделі ковзної середньої величина, що моделюється, задається лінійною функцією від збурень (помилки) у попередні моменти часу. Модель ковзної середньої q -го порядку має вигляд:

$$y_t = \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \gamma_2 \varepsilon_{t-2} + \dots + \gamma_q \varepsilon_{t-q}. \quad (4.7)$$

Іноді використовують також комбіновані моделі часових рядів. Авторегресійна модель ковзної середньої порядків p і q відповідно має вигляд

$$y_t = \beta_0 + \beta_1 y_{t-1} + \dots + \beta_p y_{t-p} + \varepsilon_t + \gamma_1 \varepsilon_{t-1} + \dots + \gamma_q \varepsilon_{t-q}. \quad (4.8)$$

Використання відповідних авторегресійних моделей для прогнозування економічних показників, тобто автопрогноз на базі розглянутих моделей, може виявитися досить ефективним (як правило, у короткостроковій перспективі).

4.5 Автокореляція залишків часового ряду.

Критерій Дарбіна-Уотсона

Автокореляція в залишках може бути викликана кількома причинами, що мають різну природу. Вона може бути пов'язана з вихідними даними і викликана наявністю помилок виміру в значеннях результативної ознаки. У ряді випад-

ків автокореляція може бути наслідком неправильної специфікації моделі. Модель може не включати фактор, що впливає на результат і вплив якого відбивається в залишках, внаслідок чого останні можуть виявитися автокорельованими. Дуже часто цим фактором є фактор часу t .

Розглянемо регресійну модель часового (динамічного) ряду. Упорядкованість спостережень виявляється істотною в тому випадку, якщо простежується механізм впливу результатів попередніх спостережень на результати наступних. Математично це виражається в тому, що випадкові величини ε_i у регресійній моделі не опиняються незалежними, зокрема, умова $r(\varepsilon_i, \varepsilon_j)=0$ не виконується. Такі моделі називають моделями з наявністю автокореляції (серіальної кореляції), на практиці ними виявляються саме часові ряди (нагадаємо, що у випадку просторової вибірки відсутність автокореляції постулюється).

Від істинної автокореляції залишків необхідно відрізнити ситуації, коли причина автокореляції полягає в неправильній специфікації функціональної форми моделі. В цьому випадку необхідно змінити форму моделі, а не використовувати спеціальні методи розрахунку параметрів рівняння регресії при наявності автокореляції в залишках.

Як правило, якщо автокореляція присутня, то найбільший вплив на наступне спостереження чинить результат попереднього спостереження. Так, наприклад, якщо розглядають ряд значень курсу якого-небудь цінного папера, то саме результат останніх торгів служить відправною точкою для формування курсу на наступних торгах.

Ситуація, коли на значення спостереження y_t впливає не результат y_{t-1} , а більш ранні значення, є досить рідкою. Частіше всього при цьому вплив носить сезонний (циклічний) характер, - наприклад, на значення y_t впливає y_{t-7} якщо спостереження здійснюються щодня і вони мають тижневий цикл (наприклад, збір кінотеатру). В цьому випадку можна скласти ряди спостережень окремо за суботами, неділями і так далі, після чого найбільш сильна кореляція буде спостерігатися між сусідніми членами.

Таким чином, відсутність кореляції між сусідніми членами служить гарною підставою вважати, що кореляція відсутня в цілому, і звичайний метод найменших квадратів дає адекватні й ефективні результати.

Один з найпоширеніших методів визначення автокореляції в залишках - це розрахунок критерію Дарбіна-Уотсона. Він визначає наявність кореляції між сусідніми членами. У критерії Дарбіна-Уотсона для оцінки кореляції використовується статистика виду

$$d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}. \quad (4.9)$$

Тобто величина d є відношенням суми квадратів різниць послідовних значень залишків до залишкової суми квадратів за моделлю регресії.

Можна показати, що при більших значеннях n існує наступне співвідношення між критерієм Дарбіна-Уотсона d і коефіцієнтом автокореляції залишків першого порядку r_1 :

$$d=2(1-r_1). \quad (4.10)$$

Таким чином, якщо в залишках існує повна додатна автокореляція і $r_1=1$, то $d=0$. Якщо в залишках повна від'ємна автокореляція, то $r_1=-1$ і, отже, $d=4$. Якщо автокореляція залишків відсутня, то $r_1=0$ і $d=2$. Таким чином, $0 \leq d \leq 4$.

Розглянемо алгоритм виявлення автокореляції залишків на основі критерію Дарбіна-Уотсона. Виставляється гіпотеза H_0 про відсутність автокореляції залишків. Альтернативні гіпотези H_1 і H_1^* складаються, відповідно, в наявності додатної або від'ємної автокореляції в залишках. Далі за спеціальними таблицями визначають критичні значення критерію Дарбіна-Уотсона d_L і d_U для заданого числа спостережень n , числа незалежних змінних моделі m і рівня значущості α . За цими значеннями числовий проміжок $[0;4]$ розбивають на п'ять відрізків. Прийняття або відхилення кожної з гіпотез з імовірністю $1-\alpha$ здійснюють в такий спосіб.

$0 < d < d_L$ – є позитивна автокореляція залишків, H_0 відхиляється, з імовірністю $p=1-\alpha$ приймається H_1 ;

$d_L < d < d_U$ – зона невизначеності;

$d_U < d < 4-d_U$ – немає підстав відхилити H_0 , тобто автокореляція залишків відсутня;

$4-d_U < d < 4-d_L$ – зона невизначеності;

$4-d_L < d < 4$ – є негативна автокореляція залишків, H_0 відхиляється, з імовірністю $p=1-\alpha$ приймається H_1^* .

Якщо фактичне значення критерію Дарбіна-Уотсона потрапляє до зони невизначеності, то на практиці припускають існування автокореляції залишків і відхиляють гіпотезу H_0 .

Існує кілька обмежень на застосування критерію Дарбіна-Уотсона. Зокрема, його неможна застосовувати до моделей, що включають як незалежні змінні лагові значення результативної ознаки. Методика розрахунку і використання критерію Дарбіна-Уотсона спрямована тільки на виявлення автокореляції залишків першого порядку. Критерій Дарбіна-Уотсона дає достовірні результати тільки для великих вибірок.

Запитання для самоперевірки

1. Наведіть визначення часового ряду.
2. Чим відрізняються часові ряди від звичайних просторових вибірок?
3. Які фактори впливають на рівні часового ряду?
4. Які компоненти містить реальний часовий ряд?
5. Поясніть аддитивну і мультиплікативну моделі часового ряду.
6. Поясніть, що таке тренд часового ряду і які види тренду зустрічаються?

7. Як визначають значущість тренду?
8. Перелічіть основні етапи аналізу часових рядів.
9. Перелічіть найпоширеніші методи аналізу часових рядів.
10. Поясніть, що таке автокореляція рівнів часового ряду. Як її можна виміряти?
11. Перелічіть властивості коефіцієнта автокореляції.
12. Що таке автокореляційна функція часового ряду і корелограма?
13. З якою метою виконують аналітичне вирівнювання часового ряду?
14. Які функції найчастіше застосовують для побудови трендів?
15. Який метод дозволяє визначити параметри тренду?
16. Які методи використовують для згладжування часового ряду?
17. В чому полягає метод ковзних середніх?
18. Поясніть, як провадиться усунення сезонних коливань за методом ковзної середньої?
19. З якою метою проводиться спектральний аналіз часового ряду?
20. Поясніть, у чому полягає зв'язний аналіз часових рядів?
21. Перелічіть основні елементи часового ряду.
22. Що являє собою автокореляція рівнів часового ряду та як її оцінюють кількісно?
23. Наведіть визначення автокореляційної функції часового ряду.
24. Перелічіть основні види трендів.
25. Яку інтерпретацію мають параметри лінійного та експоненційного трендів?
26. Запишіть загальний вигляд мультиплікативної та аддитивної моделей часового ряду.
27. Охарактеризуйте основні етапи побудови мультиплікативної та аддитивної моделей часового ряду.
28. З якою метою виявляють наявність сезонного ефекту?

Тема 5 СИСТЕМА СТРУКТУРНИХ РІВНЯНЬ

5.1 Поняття системи структурних рівнянь

При використанні окремих рівнянь регресії, наприклад для економічних розрахунків, в більшості випадків передбачають, що аргументи (фактори) можна змінювати незалежно один від одного. Однак це припущення є дуже грубим: практично зміна однієї змінної, як правило, не може відбуватися при абсолютній незмінності інших. Її зміна спричинить зміни у всій системі взаємозалежних ознак. Отже, окремо взяте рівняння множинної регресії не може характеризувати істинні впливи окремих ознак на варіацію результативної ознаки. Саме тому в останні десятиліття в економічних дослідженнях важливе місце займає проблема опису структури зв'язків між змінними системою так званих **одночасних рівнянь**, називаних також **структурними рівняннями**.

Система рівнянь в економетричних дослідженнях може бути побудована по-різному.

Можлива система незалежних рівнянь, коли кожна залежну змінну у розглядають як функцію того самого набору факторів x :

[illegible]

Набір факторів x_j в кожному рівнянні може варіювати. Кожне рівняння системи незалежних рівнянь може розглядатися самостійно. Для знаходження його параметрів використовують метод найменших квадратів. За сутністю, кожне рівняння цієї системи є рівнянням регресії. Оскільки фактичні значення залежної змінної відрізняються від теоретичних на величину випадкової помилки, то в кожному рівнянні присутня величина випадкової помилки ε_i .

Якщо залежна змінна у одного рівняння виступає у вигляді фактору x в іншому рівнянні, то дослідник може будувати модель у вигляді системи рекурсивних рівнянь:

$$\left\{ \begin{array}{l} y_1 = a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n + \varepsilon_2, \\ y_3 = b_{31}y_1 + b_{32}y_2 + a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n + \varepsilon_2, \\ \\ y_m = b_{m1}y_1 + ... + b_{m,m-1}y_{m-1} + a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n + \varepsilon_m. \end{array} \right. \quad (5.2)$$

У даній системі залежна змінна y включає в кожне наступне рівняння як фактори всі залежні змінні попередніх рівнянь поряд з набором власне факторів x . Кожне рівняння цієї системи може розглядатися самостійно, і його параметри визначаються методом найменших квадратів (МНК).

Найбільше поширення в економетричних дослідженнях одержала система взаємозалежних рівнянь. У ній ті самі залежні змінні в одних рівняннях входять до лівої частини, а в інших рівняннях - до правої частини системи:

[illegible]

Система взаємозалежних рівнянь одержала назву системи спільних, одночасних рівнянь. Тим самим підкреслюють, що в системі ті самі змінні одночасно розглядають як залежні в одних рівняннях і як незалежні в інших. В економіці цю систему рівнянь називають також **структурною формою моделі**. На відміну від попередніх систем кожне рівняння системи одночасних рівнянь

не може розглядатися самостійно, і для знаходження його параметрів традиційний МНК не можна застосовувати. З цією метою використовують спеціальні прийоми оцінювання.

5.2 Структурна і приведена форми моделі

Система спільних, одночасних рівнянь (або структурна форма моделі) зазвичай містить ендогенні і екзогенні змінні.

Ендогенні змінні – це залежні змінні, число яких дорівнює числу рівнянь у системі і які позначаються через y .

Екзогенні змінні – це певні змінні, що впливають на ендogenousні змінні, але не залежать від них, позначаються через x .

Класифікація змінних на ендогенні і екзогенні залежить від теоретичної концепції прийнятої моделі. Економічні змінні можуть виступати в одних моделях як ендогенні, а в інших - як екзогенні змінні. Позаекономічні змінні (наприклад, кліматичні умови, соціальний стан, стать, вікова категорія) входять до системи тільки як екзогенні змінні. Як екзогенні змінні можуть розглядатися значення ендогенних змінних за попередній період часу (лагові змінні).

Структурна форма моделі дозволяє побачити вплив змін будь-якої екзогенної змінної на значення ендогенної змінної. Доцільно як екзогенні змінні вибирати такі, які можуть бути об'єктом регулювання. Змінюючи їх та управляючи ними, можна заздалегідь мати цільові значення ендогенних змінних.

Структурна форма моделі в правій частині містить при ендогенних змінних коефіцієнти b_{ik} , а при екзогенних змінних – коефіцієнти a_{ij} , які називають **структурними коефіцієнтами** моделі. Всі змінні в моделі виражені у відхиленнях від середнього рівня, тобто під x мається на увазі $x - \bar{x}$, а під y – відповідно $y - \bar{y}$. Тому вільний член у кожному рівнянні системи (5.3) відсутній.

Використання МНК для оцінювання структурних коефіцієнтів моделі дає, як прийнято вважати в теорії, зміщені і неспроможні оцінки. Тому, зазвичай, для визначення структурних коефіцієнтів моделі її структурну форму перетворюють в **приведену форму** моделі.

Приведена форма моделі являє собою систему лінійних функцій ендогенних змінних від екзогенних:

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + ... + \delta_{1n}x_n + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + ... + \delta_{2n}x_n + u_2, \\ \\ y_m = \delta_{m1}x_1 + \delta_{m2}x_2 + ... + \delta_{mn}x_n + u_m, \end{cases} \quad (5.4)$$

де δ_{ij} – коефіцієнти приведеної форми моделі;

u_i – залишкова величина для приведеної форми.

За своїм виглядом приведена форма моделі нічим не відрізняється від системи незалежних рівнянь, параметри якої оцінюються традиційним МНК. За-

стосовуючи МНК, можна оцінити δ_{ij} , а потім оцінити значення ендогенних змінних через екзогенні.

Коефіцієнти приведеної форми моделі являють собою нелінійні функції коефіцієнтів структурної форми моделі. Розглянемо це положення на прикладі найпростішої структурної моделі, виразивши коефіцієнти приведеної форми моделі через коефіцієнти структурної моделі.

Для структурної моделі виду

$$\begin{cases} y_1 = b_{12}y_2 + a_{11}x_1 + \varepsilon_1, \\ y_2 = b_{21}y_1 + a_{22}x_2 + \varepsilon_2 \end{cases} \quad (5.5)$$

приведена форма моделі має вигляд

$$\begin{cases} y_1 = \delta_{11}x_1 + \delta_{12}x_2 + u_1, \\ y_2 = \delta_{21}x_1 + \delta_{22}x_2 + u_2. \end{cases} \quad (5.6)$$

З першого рівняння (5.5) можна виразити y_2 у такий спосіб (з метою спрощення опускаємо випадкову величину):

$$y_2 = \frac{y_1 + a_{11}x_1}{b_{12}}.$$

Підставляючи до другого рівняння (5.5), маємо

$$y_2 = \frac{y_1 + a_{11}x_1}{b_{12}} = b_{21}y_1 + a_{22}x_2,$$

звідки

$$y_1 = \frac{a_{11}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}} x_2.$$

Аналогічно, з другого рівняння системи (5.5), одержимо

$$y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}}{1 - b_{12}b_{21}} x_2,$$

тобто система (5.5) приймає вигляд

$$\begin{cases} y_1 = \frac{a_{11}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}b_{12}}{1 - b_{12}b_{21}} x_2, \\ y_2 = \frac{a_{11}b_{21}}{1 - b_{12}b_{21}} x_1 + \frac{a_{22}}{1 - b_{12}b_{21}} x_2. \end{cases}$$

Таким чином, можна зробити висновок, що коефіцієнти приведеної форми моделі будуть виражатися через коефіцієнти структурної форми в такий спосіб:

$$\begin{aligned} \delta_{11} &= \frac{a_{11}}{1 - b_{12}b_{21}}, & \delta_{12} &= \frac{a_{22}b_{12}}{1 - b_{12}b_{21}}, \\ \delta_{21} &= \frac{a_{11}b_{21}}{1 - b_{12}b_{21}}, & \delta_{22} &= \frac{a_{22}}{1 - b_{12}b_{21}}. \end{aligned}$$

Треба помітити, що приведена форма моделі, хоча й дозволяє одержати значення ендогенної змінної через значення екзогенних змінних, але аналітично вона уступає структурній формі моделі, тому що в ній відсутні оцінки взаємозв'язку між ендогенними змінними.

5.3 Проблема ідентифікації

При переході від приведеної форми моделі до структурної економетрист зіштовхується з проблемою ідентифікації. Ідентифікація - це однозначність відповідності між приведеною і структурною формами моделі.

Структурна модель (5.3) у повному вигляді містить $m(m+n-1)$ параметрів, а приведена форма моделі в повному вигляді містить mn параметрів. Таким чином, у повному вигляді структурна модель містить більшу кількість параметрів, ніж приведена форма моделі. Відповідно $m(m+n-1)$ параметрів структурної моделі не можуть бути однозначно визначені з mn параметрів приведеної форми моделі.

Щоб одержати єдино можливе рішення для структурної моделі, необхідно припустити, що деякі з структурних коефіцієнтів у зв'язку із слабким взаємозв'язком ознак з ендогенною змінною з лівої частини системи дорівнюють нулю. В такий спосіб зменшиться число структурних коефіцієнтів моделі. Зменшення числа структурних коефіцієнтів моделі можливе і другим шляхом, наприклад, шляхом прирівнювання деяких коефіцієнтів один до одного, тобто шляхом припущень, що їхній вплив на формовану ендогенну змінну однаковий. На структурні коефіцієнти можуть накладатися, наприклад, обмеження виду $b_{ik} + a_{ij} = 0$.

З позиції ідентифікованості структурні моделі можна підрозділити на три види:

- ідентифіковані;
- неідентифіковані;
- надідентифіковані.

Модель є **ідентифікованою**, якщо всі її структурні коефіцієнти визначаються однозначно, єдиним образом за коефіцієнтами приведеної форми моделі, тобто якщо число параметрів структурної моделі дорівнює числу параметрів приведеної форми моделі. В цьому випадку структурні коефіцієнти моделі оцінюються через параметри приведеної форми моделі і модель є ідентифікованою.

Модель **неідентифікована**, якщо число наведених коефіцієнтів менше числа структурних коефіцієнтів, і в результаті структурні коефіцієнти не можуть бути оцінені через коефіцієнти приведеної форми моделі.

Модель **надідентифікована**, якщо число наведених коефіцієнтів більше числа структурних коефіцієнтів. В цьому випадку на основі коефіцієнтів приведеної форми можна одержати два або більше значення одного структурного коефіцієнта. В цієї моделі число структурних коефіцієнтів менше числа коефіцієнтів приведеної форми. Надідентифікована модель на відміну від неідентифіко-

ваної моделі практично розв'язувана, але вимагає для цього спеціальних методів вирахування параметрів.

Структурна модель завжди являє собою систему спільних рівнянь, кожне з яких потрібно перевіряти на ідентифікацію. Модель вважається ідентифікованою, якщо кожне рівняння системи є ідентифікованим. Якщо хоча б одне з рівнянь системи неідентифіковане, то і вся модель вважається неідентифікованою. Надідентифікована модель містить хоча б одне надідентифіковане рівняння.

Виконання умови ідентифікованості моделі перевіряється для кожного рівняння системи. Щоб рівняння було ідентифіковане, необхідно, щоб число визначених змінних, відсутніх у даному рівнянні, але присутніх у системі, дорівнювало числу ендогенних змінних в даному рівнянні без одного.

Якщо позначити H число ендогенних змінних в i -му рівнянні системи, а число екзогенних змінних, які містяться в системі, але не входять в дане рівняння - D , то умова ідентифікованості моделі може бути записана у вигляді наступного рахункового правила:

$$\begin{array}{ll} D+1=H & \text{рівняння ідентифіковане,} \\ D+1<H & \text{рівняння неідентифіковане,} \\ D+1>H & \text{рівняння надідентифіковане.} \end{array}$$

Для оцінки параметрів структурної моделі система має бути ідентифікованою або надідентифікованою.

Розглянуте рахункове правило відбиває необхідну, але недостатню умову ідентифікації. Більш точно умови ідентифікації визначаються, якщо накладати обмеження на коефіцієнти матриць параметрів структурної моделі. Рівняння ідентифіковане, якщо по відсутнім у ньому змінним (ендогенним і екзогенним) можна з коефіцієнтів при них в інших рівняннях системи одержати матрицю, визначник якої не дорівнює нулю, а ранг матриці не менший, ніж число ендогенних змінних у системі без одного.

Доцільність перевірки умови ідентифікації моделі через визначник матриці коефіцієнтів, відсутніх у даному рівнянні, але присутніх в інших, пояснюється тим, що можлива ситуація, коли для кожного рівняння системи виконане рахункове правило, а визначник матриці названих коефіцієнтів дорівнює нулю. В цьому випадку дотримується лише необхідна, але недостатня умова ідентифікації.

В економетричних моделях часто поряд з рівняннями, параметри яких повинні бути статистично оцінені, використовують балансові тотожності змінних, коефіцієнти при яких дорівнюють ± 1 . У цьому випадку, хоча сама тотожність і не вимагає перевірки на ідентифікацію, тому що коефіцієнти при змінних у тотожності відомі, у перевірці на ідентифікацію властиво структурних рівнянь системи тотожності беруть участь.

Розглянемо приклад. Вивчають модель виду

$$\begin{cases} C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1, \\ I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2, \\ r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3, \\ Y_t = C_t + I_t + G_t, \end{cases}$$

де C_t – витрати на споживання в період t ;

Y_t – сукупний дохід у період t ;

I_t – інвестиції в період t ;

r_t – процентна ставка в період t ;

M_t – грошова маса в період t ;

G_t – державні витрати в період t ;

C_{t-1} – витрати на споживання в період $t-1$;

I_{t-1} – інвестиції в період $t-1$.

Перше рівняння – функція споживання, друге рівняння – функція інвестицій, третє рівняння – функція грошового ринку, четверте рівняння – тотожність доходу.

Модель являє собою систему одночасних рівнянь. Перевіримо кожне її рівняння на ідентифікацію.

Модель включає чотири ендогенні змінні (C_t , I_t , Y_t , r_t) і чотири визначені змінні (дві екзогенні змінні – M_t і G_t і дві лагові змінні – C_{t-1} і I_{t-1}).

Перевіримо необхідну умову ідентифікації для кожного з рівнянь моделі.

Перше рівняння: $C_t = a_1 + b_{11}Y_t + b_{12}C_{t-1} + \varepsilon_1$. Це рівняння містить дві ендогенні змінні C_t і Y_t і одну визначену змінну C_{t-1} . Таким чином, $H=2$, а $D=4-1=3$, тобто виконується умова $D+1>H$. Рівняння надідентифіковане.

Друге рівняння: $I_t = a_2 + b_{21}r_t + b_{22}I_{t-1} + \varepsilon_2$. Воно включає дві ендогенні змінні I_t і r_t і одну екзогенну змінну I_{t-1} . Виконується умова $D+1=3+1>H=2$. Рівняння надідентифіковане.

Третє рівняння: $r_t = a_3 + b_{31}Y_t + b_{32}M_t + \varepsilon_3$. Воно включає дві ендогенні змінні Y_t і r_t і одну екзогенну змінну M_t . Виконується умова $D+1=3+1>H=2$. Рівняння надідентифіковане.

Четверте рівняння: $Y_t = C_t + I_t + G_t$. Воно являє собою тотожність, параметри якої відомі. Необхідності в ідентифікації немає.

Перевіримо для кожного рівняння достатність умови ідентифікації. Для цього складемо матрицю коефіцієнтів при змінних моделі.

	C_t	I_t	r_t	Y_t	C_{t-1}	I_{t-1}	M_t	G_t
I рівняння	-1	0	0	b_{11}	b_{12}	0	0	0
II рівняння	0	-1	b_{21}	0	0	b_{22}	0	0
III рівняння	0	0	-1	b_{31}	0	0	b_{32}	0
Тотожність	1	1	0	-1	0	0	0	1

Відповідно до достатньої умови ідентифікації ранг матриці коефіцієнтів при змінних, що не входять до досліджуваного рівняння, повинен дорівнювати числу ендогенних змінних моделі без одної.

Перше рівняння. Матриця коефіцієнтів при змінних, що не входять у рівняння, має вигляд

$$\begin{array}{l} \text{II рівняння} \\ \text{III рівняння} \\ \text{Тотожність} \end{array} \begin{vmatrix} I_t & r_t & I_{t-1} & M_t & G_t \\ -1 & b_{21} & b_{22} & 0 & 0 \\ 0 & -1 & 0 & b_{32} & 0 \\ 1 & 0 & 0 & 0 & 1 \end{vmatrix}$$

Ранг даної матриці дорівнює трьом, оскільки визначник квадратної підматриці 3×3 не дорівнює нулю:

$$\begin{vmatrix} b_{22} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{22}b_{32} \neq 0.$$

Достатня умова ідентифікації для даного рівняння виконується.

Друге рівняння. Матриця коефіцієнтів при змінних, що не входять до рівняння, має вигляд

$$\begin{array}{l} \text{I рівняння} \\ \text{III рівняння} \\ \text{Тотожність} \end{array} \begin{vmatrix} C_t & Y_t & C_{t-1} & M_t & G_t \\ -1 & b_{11} & b_{12} & 0 & 0 \\ 0 & b_{31} & 0 & b_{32} & 0 \\ 1 & -1 & 0 & 0 & 1 \end{vmatrix}$$

Ранг даної матриці дорівнює трьом, оскільки визначник квадратної підматриці 3×3 не дорівнює нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{32} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{32} \neq 0.$$

Достатня умова ідентифікації для даного рівняння виконується.

Третє рівняння. Матриця коефіцієнтів при змінних, що не входять до рівняння, має вигляд

$$\begin{array}{l} \text{I рівняння} \\ \text{II рівняння} \\ \text{Тотожність} \end{array} \begin{vmatrix} C_t & I_t & C_{t-1} & I_{t-1} & G_t \\ -1 & 0 & b_{12} & 0 & 0 \\ 0 & -1 & 0 & b_{22} & 0 \\ 1 & 1 & 0 & 0 & 1 \end{vmatrix}$$

Ранг даної матриці дорівнює трьом, оскільки визначник квадратної підматриці 3×3 не дорівнює нулю:

$$\begin{vmatrix} b_{12} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 1 \end{vmatrix} = b_{12}b_{22} \neq 0.$$

Достатня умова ідентифікації для даного рівняння виконується.

Таким чином, всі рівняння моделі надідентифіковані. Приведена форма моделі в загальному виді матиме такий вигляд:

$$\begin{cases} C_t = A_1 + \delta_{11}C_{t-1} + \delta_{12}I_{t-1} + \delta_{13}M_t + \delta_{14}G_t + u_1, \\ I_t = A_2 + \delta_{21}C_{t-1} + \delta_{22}I_{t-1} + \delta_{23}M_t + \delta_{24}G_t + u_2, \\ r_t = A_3 + \delta_{31}C_{t-1} + \delta_{32}I_{t-1} + \delta_{33}M_t + \delta_{34}G_t + u_3, \\ Y_t = A_4 + \delta_{41}C_{t-1} + \delta_{42}I_{t-1} + \delta_{43}M_t + \delta_{44}G_t + u_4. \end{cases}$$

5.4 Методи оцінки параметрів структурної форми моделі

Коефіцієнти структурної моделі можуть бути оцінені різними способами залежно від виду системи одночасних рівнянь. Найбільшого поширення в літературі одержали наступні методи оцінювання коефіцієнтів структурної моделі:

- непрямий метод найменших квадратів;
- двокроковий метод найменших квадратів;
- трикроковий метод найменших квадратів;
- метод максимальної правдоподібності з повною інформацією;
- метод максимальної правдоподібності при обмеженій інформації.

Розглянемо сутність кожного з цих методів.

Непрямий метод найменших квадратів (НМНК) застосовують у випадку точно ідентифікованої структурної моделі. Процедура застосування НМНК припускає виконання наступних етапів роботи:

- структурна модель перетворюється в приведену форму моделі;
- для кожного рівняння приведеної форми моделі за звичайним МНК оцінюються приведені коефіцієнти δ_{ij} ;
- коефіцієнти приведеної форми моделі трансформуються в параметри структурної моделі.

Якщо система надідентифікована, то НМНК не використовують, тому що він не дає однозначних оцінок для параметрів структурної моделі. В цьому випадку можуть використовуватися різні методи оцінювання, серед яких найпоширенішим і простим є двокроковий метод найменших квадратів (ДМНК).

Основна ідея ДМНК - на основі приведеної форми моделі одержати для надідентифікованого рівняння теоретичні значення ендогенних змінних, що містяться в правій частині рівняння. Далі, підставивши їх замість фактичних значень, можна застосувати звичайний МНК до структурної форми надідентифікованого рівняння. Метод одержав назву двокрокового МНК, тому що двічі використовує МНК: на першому кроці при визначенні приведеної форми моделі і знаходженні на її основі оцінок теоретичних значень ендогенної змінної

$\hat{y}_i = \delta_{i1}x_1 + \delta_{i2}x_2 + \dots + \delta_{in}x_n$ і на другому кроці стосовно до структурного надідентифікованого рівняння при визначенні структурних коефіцієнтів моделі за даними теоретичних (розрахункових) значень ендогенних змінних.

Надідентифікована структурна модель може бути двох типів:

- всі рівняння системи є надідентифікованими;
- система містить поряд з надідентифікованими точно ідентифіковані рівняння.

Якщо всі рівняння системи є надідентифіковані, то для оцінки структурних коефіцієнтів кожного рівняння використовують ДМНК. Якщо в системі є точно ідентифіковані рівняння, то структурні коефіцієнти за ними визначаються з системи приведених рівнянь.

Для прикладу, розглянутого в попередньому підрозділі, необхідно застосувати саме двокроковий метод найменших квадратів. Але можна зробити наступне зауваження. Якщо з моделі виключити тотожність доходу, число ендогенних змінних моделі знизиться на одиницю – змінна Y_t стане екзогенною. Число визначених змінних моделі не зміниться, тому що з моделі буде виключена ендогенна змінна G_t , але її місце займе змінна Y_t . В правих частинах функції споживання і функції грошового ринку будуть перебувати тільки визначені змінні. Функція інвестицій постулює залежність ендогенної змінної I_t від ендогенної змінної r_t , (яка залежить тільки від наперед визначених змінних) і визначеної змінної I_{t-1} . Таким чином, ми одержимо рекурсивну систему. Її параметри можна оцінювати звичайним МНК, і немає необхідності досліджувати рівняння на ідентифікацію.

Непрямий і двокроковий методи найменших квадратів докладно описані в літературі і розглядаються як традиційні методи оцінки коефіцієнтів структурної моделі. Ці методи досить легко реалізовані.

Метод максимальної правдоподібності розглядають як найбільш загальний метод оцінювання, результати якого при нормальному розподілі ознак збігаються з МНК. Однак при великій кількості рівнянь системи цей метод приводить до досить складних обчислювальних процедур. Тому як модифікацію використовують метод максимальної правдоподібності при обмеженій інформації (метод найменшого дисперсійного відношення), розроблений в 1949 р. Т. Андерсоном і Н. Рубіним.

На відміну від методу максимальної правдоподібності в даному методі зняті обмеження на параметри, що пов'язані з функціонуванням системи в цілому. Це робить розв'язання більш простим, але трудомісткість обчислень залишається досить високою. Незважаючи на його значну популярність, до середини 60-х років ХХ ст. він був практично витиснутий двокроковим методом найменших квадратів (ДМНК) у зв'язку з набагато більшою простотою останнього.

Подальшим розвитком ДМНК є трикроковий МНК (ТМНК), запропонований у 1962 р. А. Зельнером і Г. Тейлом. Цей метод оцінювання придатний для всіх видів рівнянь структурної моделі. Однак, при деяких обмеженнях на параметри ефективнішим виявляється ДМНК.

Запитання для самоперевірки

1. Які основні причини використання систем одночасних рівнянь?
2. Перелічіть можливі способи побудови систем рівнянь. Чим вони відрізняються один від одного?
3. Як пов'язані одна з одною структурна та приведена форми моделі? У чому їх основна розбіжність?
4. Чому для оцінювання систем одночасних рівнянь майже не використовують звичайний метод найменших квадратів?
5. У чому полягає сутність непрямого методу найменших квадратів?
6. У чому полягають проблеми ідентифікації моделі? Які умови ідентифікації існують?
7. Охарактеризуйте причини неідентифікованості та надідентифікованості систем структурних рівнянь?

ЗМ 3 МЕТОДИ АНАЛІЗУ НА ПІДСТАВІ СТАТИСТИЧНИХ РІВНЯНЬ. МОДЕЛЬ З АВТОКОРЕЛЬОВАНИМИ ЗАЛИШКАМИ. МОДЕЛІ РОЗПОДІЛЕНОГО ЛАГУ

Тема 6 НЕЛІНІЙНІ ОДНОФАКТОРНІ МОДЕЛІ

6.1 Види нелінійних моделей парної регресії

Якщо між економічними явищами існують нелінійні співвідношення, то вони виражаються за допомогою відповідних нелінійних функцій. Дотепер ми розглядали лінійні регресійні моделі, в яких змінні мали першу степінь (моделі, лінійні за змінними), а параметри виступали у вигляді коефіцієнтів при цих змінних (моделі, лінійні за параметрами). Проте співвідношення між соціально-економічними явищами і процесами далеко не завжди можна виразити лінійними функціями, тому що при цьому можуть виникати невиправдано великі помилки.

Наприклад, нелінійними виявляються виробничі функції (залежності між обсягом виробленої продукції і основними факторами виробництва – працею, капіталом та ін.), функції попиту (залежність між попитом на товари або послуги і їхні ціни, або доходом) та ін.

Для оцінки параметрів нелінійних моделей використовують два підходи. Перший підхід заснований на лінеаризації моделі і полягає в тому, що за допомогою перетворень вихідних змінних залежність, яку досліджують, представляють у вигляді лінійного співвідношення між перетвореними змінними.

Другий підхід зазвичай застосовують у випадку, коли підібрати відповідне перетворення щодо лінеаризації не вдається. В цьому випадку застосовують методи нелінійної оптимізації на основі вихідних змінних.

Для лінеаризації моделі в рамках першого підходу можуть використовуватися як моделі, нелінійні за змінними, так і нелінійні за параметрами.

Якщо модель нелінійна за змінними, то введенням нових змінних її можна звести до лінійної моделі, для оцінки параметрів якої скористатися звичайним методом найменших квадратів. Проте потрібно відзначити й недолік такої заміни змінних, пов'язаний з тим, що оцінки параметрів виходять не з умови мінімізації суми квадратів відхилень для вихідних змінних, а з умови мінімізації суми квадратів відхилень для перетворених змінних, що не те саме. У зв'язку з цим необхідне певне уточнення отриманих оцінок.

Більш складною проблемою є нелінійність моделі за параметрами, тому що безпосереднє застосування методу найменших квадратів для їхнього оцінювання неможливо. До числа таких моделей можна віднести, наприклад, мультиплікативну (ступеневу) модель

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} \varepsilon_i, i=1,n; \quad (6.1)$$

експонентну модель

$$y_i = e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}} \varepsilon_i, \quad i=1, n, \quad (6.2)$$

та інші. У ряді випадків шляхом перетворень ці моделі вдається привести до лінійної форми. Так, моделі (6.1) і (6.2) можуть бути приведені до лінійних логарифмуванням обох частин рівнянь. Тоді, наприклад, модель (6.1) прийме вид:

$$\ln y_i = \ln \beta_0 + \beta_1 \ln x_{i1} + \beta_2 x_{i2} + \ln \varepsilon_i, \quad i=1, n. \quad (6.3)$$

До моделі (6.3) уже можна застосовувати звичайні методи дослідження лінійної регресії. Проте треба підкреслити, що критерії значущості та інтервальні оцінки параметрів, які застосовують для нормальної лінійної регресії, вимагають, щоб нормальний закон розподілу в моделях (6.1), (6.2) мав логарифм вектору збурень ε , а не \mathcal{E} . Інакше кажучи, вектор збурень \mathcal{E} повинен мати логарифмічно нормальний розподіл.

Помітимо також, що до моделі

$$y_i = \beta_0 x_{i1}^{\beta_1} x_{i2}^{\beta_2} + \varepsilon_i, \quad i=1, n, \quad (6.4)$$

яку розглянуто як альтернативну щодо моделі (6.1), викладені вище методи дослідження лінійної регресії вже непридатні, тому що модель (6.4) не можна привести до лінійного виду. В цьому випадку використовують спеціальні (ітеративні) процедури оцінювання параметрів.

Розглянемо два класи нелінійних регресій:

1. Регресії, нелінійні щодо включених до аналізу пояснюючих змінних, але лінійні за оцінюваними параметрами, наприклад

– поліноми різних ступенів – $\hat{y}_x = a + bx + cx^2$, $\hat{y}_x = a + bx + cx^2 + dx^3$;

– рівнобічна гіпербола – $\hat{y}_x = a + \frac{b}{x}$;

– напівлогарифмічна функція – $\hat{y}_x = a + b \ln x$.

2. Регресії, нелінійні за оцінюваними параметрами, наприклад

– ступенева – $\hat{y}_x = ax^b$;

– показова – $\hat{y}_x = a \cdot b^x$;

– експонентна – $\hat{y}_x = e^{a+bx}$.

6.2 Моделі нелінійні щодо пояснюючих змінних

Регресії, нелінійні за включеними змінними, приводять до лінійного виду простою заміною змінних, а подальшу оцінку параметрів здійснюють за допомогою методу найменших квадратів. Розглянемо деякі функції.

Поліном другого ступеня $\hat{y}_x = a + bx + cx^2$ приводиться до лінійного виду за допомогою заміни: $x = x_1$, $x^2 = x_2$. В результаті приходимо до двофакторного рівняння $\hat{y}_x = a + bx_1 + cx_2$, оцінка параметрів якого за допомогою МНК, як буде показано далі, призводить до системи наступних нормальних рівнянь:

$$\begin{cases} an + b \sum_{i=1}^n x_{1i} + c \sum_{i=1}^n x_{2i} = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_{1i} + b \sum_{i=1}^n x_{1i}^2 + c \sum_{i=1}^n x_{1i} x_{2i} = \sum_{i=1}^n x_{1i} y_i, \\ a \sum_{i=1}^n x_{2i} + b \sum_{i=1}^n x_{1i} x_{2i} + c \sum_{i=1}^n x_{2i}^2 = \sum_{i=1}^n x_{2i} y_i, \end{cases}$$

з якої після зворотної заміни змінних одержимо

$$\begin{cases} an + b \sum_{i=1}^n x_i + c \sum_{i=1}^n x_i^2 = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 + c \sum_{i=1}^n x_i^3 = \sum_{i=1}^n x_i y_i, \\ a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i^3 + c \sum_{i=1}^n x_i^4 = \sum_{i=1}^n x_i^2 y_i. \end{cases} \quad (6.5)$$

Поліном другого ступеня зазвичай застосовують у випадках, коли для певного інтервалу значень фактору змінюється характер зв'язку розглянутих ознак: прямий зв'язок змінюється на зворотний або зворотний на прямий.

Рівнобічна гіпербола $\hat{y}_x = a + \frac{b}{x}$ може бути використана для характеристики зв'язку питомих витрат сировини, матеріалів, палива та обсягу продукції, часу обігу товарів і величини товарообігу, відсотку приросту заробітної плати та рівня безробіття (наприклад, крива А. В. Філіпса), витрат на непродуктивні товари і доходів або загальної суми витрат (наприклад, криві Е. Енгеля) та в інших випадках. Гіпербола приводиться до лінійного рівняння простою заміною: $z = \frac{1}{x}$. Система лінійних рівнянь при використанні МНК буде наступною:

$$\begin{cases} an + b \sum_{i=1}^n \frac{1}{x_i} = \sum_{i=1}^n y_i, \\ a \sum_{i=1}^n \frac{1}{x_i} + b \sum_{i=1}^n \frac{1}{x_i^2} = \sum_{i=1}^n \frac{1}{x_i} y_i. \end{cases} \quad (6.6)$$

Аналогічно приводять до лінійного виду залежності $\hat{y}_x = a + b \ln x$, $\hat{y}_x = a + b\sqrt{x}$ та інші.

6.3 Моделі нелінійні за оцінюваними параметрами

Дещо інший вигляд має справа з регресіями, які є нелінійними за оцінюваними параметрами. Вони в свою чергу діляться на два типи: нелінійні моделі внутрішньолінійні (приводяться до лінійного виду за допомогою відповідних перетворень, наприклад, логарифмуванням) і нелінійні моделі внутрішньонелі-

нійні (до лінійного виду не приводяться).

До внутрішньолінійних моделей належать, наприклад, степенева функція – $\hat{y}_x = ax^b$, показова – $\hat{y}_x = a \cdot b^x$, експонентна – $\hat{y}_x = e^{a+bx}$, логістична – $\hat{y}_x = \frac{a}{1+be^{-cx}}$, зворотна – $\hat{y}_x = \frac{1}{a+bx}$.

До внутрішньонелінійних моделей можна, наприклад, віднести наступні моделі: $\hat{y}_x = a + bx^c$, $\hat{y}_x = a \left(1 - \frac{1}{1-x^b} \right)$.

Серед нелінійних моделей найбільш часто використовують степеневу функцію $y = ax^b \varepsilon$, яку приводять до лінійного виду логарифмуванням:

$$\ln y = \ln a + b \ln x + \ln \varepsilon ;$$

$$Y = A + bX + E ,$$

де $Y=\ln y$, $X=\ln x$, $A=\ln a$, $E=\ln \varepsilon$.

Потім використовують МНК для визначення невідомих параметрів A і b :

$$\begin{cases} An + b \sum_{i=1}^n X_i = \sum_{i=1}^n Y_i, \\ A \sum_{i=1}^n X_i + b \sum_{i=1}^n X_i^2 = \sum_{i=1}^n X_i Y_i \end{cases}$$

і потенціюванням одержують шукане рівняння.

Широке використання степеневі функції пов'язане з тим, що параметр b в ній має чітке економічне тлумачення - він є коефіцієнтом еластичності. (Коефіцієнт еластичності показує, на скільки відсотків зміниться в середньому результативна ознака, якщо фактор зміниться на 1%). Формула для розрахунку коефіцієнту еластичності має вигляд:

$$E = f'(x) \frac{x}{y} . \quad (6.7)$$

Оскільки для інших функцій коефіцієнт еластичності не є постійною величиною, а залежить від відповідного значення фактору x , то зазвичай розраховують середній коефіцієнт еластичності:

$$E = f'(\bar{x}) \frac{\bar{x}}{\bar{y}} . \quad (6.8)$$

Можливі випадки, коли розрахунок коефіцієнта еластичності не має смислу. Це відбувається тоді, коли для розглянутих ознак немає смислу визначати зміни у відсотках.

Рівняння нелінійної регресії, так само, як і у випадку лінійної залежності, доповнюють показником тісноти зв'язку. У цьому випадку це індекс кореляції:

$$\rho_{xy} = \sqrt{1 - \frac{\sigma_{\text{зал}}^2}{\sigma_y^2}}, \quad (6.9)$$

де $\sigma_y^2 = \frac{1}{n} \sum_{i=1}^n (y - \bar{y})^2$ – загальна дисперсія результативної ознаки y ;

$\sigma_{\text{зал}}^2 = \frac{1}{n} \sum_{i=1}^n (y - \hat{y}_x)^2$ – залишкова дисперсія.

Величина даного показника знаходиться в межах: $0 \leq \rho_{xy} \leq 1$. Чим ближче значення індексу кореляції до одиниці, тим тісніше зв'язок розглянутих ознак і більш надійне рівняння регресії.

Квадрат індексу кореляції називають індексом детермінації, він характеризує частку дисперсії результативної ознаки y , що пояснюється регресією, у загальній дисперсії результативної ознаки:

$$\rho_{xy}^2 = 1 - \frac{\sigma_{\text{ост}}^2}{\sigma_y^2} = \frac{\sigma_{\text{об}}^2}{\sigma_y^2}, \quad (6.10)$$

тобто має той самий зміст, що і коефіцієнт детермінації в лінійній регресії:

$$\sigma_{\text{об}}^2 = \frac{1}{n} \sum_{i=1}^n (\hat{y}_x - \bar{y})^2.$$

Індекс детермінації ρ_{xy}^2 можна порівнювати з коефіцієнтом детермінації r_{xy}^2 за розміром для обґрунтування можливості застосування лінійної функції. Чим більше кривизна лінії регресії, тим розмір r_{xy}^2 менше ρ_{xy}^2 . А близькість цих показників вказує на те, що немає необхідності ускладнювати форму рівняння регресії і можна використати лінійну функцію.

Індекс детермінації використовують для перевірки істотності в цілому рівняння регресії за F -критерієм Фішера:

$$F = \frac{\rho_{xy}^2}{1 - \rho_{xy}^2} * \frac{n - m - 1}{m}, \quad (6.11)$$

де ρ_{xy}^2 – індекс детермінації;

n – число спостережень;

m – число параметрів при змінній x .

Фактичне значення F -критерію (6.11) порівнюють з табличним при рівні значущості α і числі ступенів свободи $k_2 = n - m - 1$ (для залишкової суми квадратів) і $k_1 = m$ (для факторної суми квадратів).

Про якість нелінійного рівняння регресії можна також судити і за середньою помилкою апроксимації, яку, так само як і в лінійному випадку, обчислюють за формулою (2.13).

6.4 Виробнича функція Кобба-Дугласа

Як приклад використання перетворення регресії, що лінеаризує, розглянемо виробничу функцію Кобба-Дугласа

$$Y = AK^{\alpha}L^{\beta}, \quad (6.12)$$

де Y - обсяг виробництва;

K - витрати капіталу;

L - витрати праці.

З огляду на вплив випадкових збурень, що властиві кожному економічному явищу, функцію Кобба-Дугласа (6.12) можна уявити у вигляді

$$Y = AK^{\alpha}L^{\beta}\varepsilon. \quad (6.13)$$

Отриману мультиплікативну (степеневу) модель легко звести до лінійної шляхом логарифмування обох частин рівняння (6.13). Тоді для i -го спостереження одержимо

$$\ln y_i = \ln A + \alpha \ln K_i + \beta \ln L_i + \ln \varepsilon_i, \quad i=1, n. \quad (6.14)$$

Якщо в моделі (6.13) $\alpha+\beta=1$ (тобто модель така, що при розширенні масштабу виробництва - збільшенні витрат капіталу K і праці L у певне число разів - обсяг виробництва зростає в те саме число разів) функцію Кобба-Дугласа представляють у вигляді

$$Y = AK^{\alpha}L^{1-\alpha}\varepsilon,$$

або

$$\frac{Y}{L} = A \left(\frac{K}{L} \right)^{\alpha} \varepsilon. \quad (6.15)$$

Таким чином, одержуємо залежність продуктивності праці $\frac{Y}{L}$ від її капіталоозброєності $\frac{K}{L}$. Для оцінки параметрів моделі (6.15) шляхом логарифмування приведемо її до вигляду (для i -го спостереження)

$$\ln \left(\frac{Y}{L} \right)_i = \ln A + \alpha \ln \left(\frac{K}{L} \right)_i + \ln \varepsilon_i, \quad i=1, n. \quad (6.16)$$

Розглянемо властивості виробничої функції Кобба-Дугласа:

- еластичність випуску продукції;
- ефект від масштабу виробництва;
- прогнозування частки виробничих факторів.

Показники α і β є коефіцієнтами часткової еластичності обсягу виробництва Y відповідно за витратами капіталу K і праці L . Це означає, що при збільшенні одних тільки витрат капіталу (праці) на 1% обсяг виробництва збільшиться на $\alpha\%$ ($\beta\%$).

Нагадаємо, що коефіцієнтом часткової еластичності $E_{x_i}(y)$ функції $y=f(x_1, x_2, \dots, x_n)$ щодо змінної x_i ($i=1, n$) називають границю відношення відносного часткового приросту функції до відносного приросту цієї змінної при $\Delta x_i \rightarrow 0$, тобто

$$E_{x_i}(y) = \lim_{\Delta x_i \rightarrow 0} \left(\frac{\Delta x_i y}{y} : \frac{\Delta x_i}{x_i} \right) = \frac{x_i}{y} y'_{x_i}.$$

Очевидно, що для функції Кобба-Дугласа

$$E_K(Y) = \alpha, \quad E_L(Y) = \beta.$$

Обидві величини α і β приймають значення між нулем і одиницею. Вони є додатними, тому що збільшення витрат виробничих факторів має викликати зростання виробництва. І вони повинні бути меншими за одиницю, тому що зниження ефекту від масштабу виробництва призводить до повільнішого росту випуску продукції, порівняно з витратами виробничих факторів, якщо інші фактори залишаються постійними.

Якщо сума α і β перевищує одиницю, то функція має зростаючий ефект від масштабу виробництва, тобто при збільшенні K і L Y збільшується в більшій мірі. Якщо сума α і β дорівнює одиниці, то ефект від масштабу виробництва постійний, тобто при збільшенні K і L Y збільшується в такій самій мірі. Якщо сума α і β менша за одиницю, то має місце спадний ефект від масштабу виробництва, тобто Y збільшується в меншій мірі, ніж K і L .

Відповідно до припущення про конкурентність ринків факторів виробництва α і β мають інтерпретацію як прогнозовані частки доходу, отриманого за рахунок капіталу та праці відповідно. Якщо ринок праці має конкурентний характер, то ставка заробітної плати w дорівнюватиме граничному продукту праці

$$w = \frac{\partial Y}{\partial L} = \alpha K^\alpha \beta L^{\beta-1} = \frac{\beta Y}{L}. \quad (6.17)$$

Отже, загальна сума заробітної плати w дорівнюватиме βY , а частка праці в загальному випуску продукції $\frac{wL}{Y}$ дорівнює постійній величині β .

Норма прибутку ρ виражається через $\frac{\partial Y}{\partial K}$

$$\rho = \frac{\partial Y}{\partial K} = \alpha \alpha K^{\alpha-1} L^\beta = \frac{\alpha Y}{K}. \quad (6.18)$$

Отже, загальний прибуток ρK дорівнює αY , а частка прибутку дорівнює постійній величині α .

Функція Кобба-Дугласа з урахуванням технічного прогресу має вигляд:

$$Y = \alpha K^\alpha L^\beta e^{\theta t} \varepsilon, \quad (6.19)$$

де t - час;

θ - темп приросту обсягу виробництва завдяки технічному прогресу. Модель (6.19) приводять до лінійного виду аналогічно моделі (6.15).

Запитання для самоперевірки

1. Якою нелінійною функцією можна замінити параболу другої степені, якщо не спостережується зміна спрямованості зв'язку ознак?
2. Перелічіть види моделей нелінійних щодо уведених змінних та оцінюваних параметрів.
3. Охарактеризуйте відмінності у застосуванні методу найменших квадратів до моделей нелінійних щодо змінних та до моделей нелінійних щодо оцінюваних параметрів.
4. Як визначають коефіцієнти еластичності за різними видами регресійних моделей?
5. Які показники кореляції, що використовують за нелінійні співвідношення досліджуваних ознак?
6. Як обирають перетворення лінеаризації для моделей з внутрішньою нелінійністю?
7. За якими ознаками класифікують методи оцінювання параметрів нелінійних моделей?

Тема 7 РЕГРЕСІЙНІ ДИНАМІЧНІ МОДЕЛІ

7.1 Причини виникнення лагових ефектів в економетричних моделях

Дотепер розглядалися моделі, в яких пояснюючі змінні x_j ($j=1, p$), не були випадковими. Це означає, що якби ми повторили серію вибірових спостережень, значення змінних x_j виявилися б тими самими, в той час як значення y змінилися б за рахунок випадкового члену ε . Таке припущення, що призводить до значних спрощень, може бути виправдано в тому випадку, коли експериментальні дані являють собою просторову вибірку. Справді, можна вважати, що значення змінних x_j обирають заздалегідь, а потім спостерігають значення y , що виходять при цьому.

У випадку часового ряду, регресори якого являють собою часовий тренд, циклічну і сезонну компоненти, пояснюючі змінні також, мабуть, не випадкові. Проте у тих випадках, коли серед регресорів часового ряду присутні змінні, значення яких самі утворюють часовий ряд, припущення щодо їхньої детермінованості неправомірне. Так що в моделях часових рядів, як правило, x_{tj} ($t=1, n; j=1, p$) спостереження мають вважатися випадковими величинами.

В цьому випадку природно виникає питання про наявність кореляції між регресорами і помилками регресії ε . Від цього істотно залежать результати оцінювання регресійної моделі, причому не тільки кількісно, але і якісно.

Можна показати, що якщо регресори x і помилки регресії ε не корелюють, то оцінка параметру b виявляється в цьому випадку незміщеною і спроможною. Якщо значення регресорів X не корельовані з помилками регресії ε у поточний момент часу t , але корелюють з помилками регресії в більш ранні моменти часу $t-\tau$, то в цьому випадку оцінка параметру b залишається спроможною, проте вона вже не буде незміщеною. Якщо значення регресорів x корельовані з помилками ε , то в цьому випадку оцінка параметру b не є ні спроможною, ні незміщеною.

Отже, якщо розглядають модель

$$y_t = a + bx_t + \varepsilon \quad (7.1)$$

із стохастичними регресорами, то оцінки параметра b , отримані за методом найменших квадратів:

- незміщені і спроможні, якщо пояснюючі змінні і помилки регресії не корелюють;
- спроможні, але зміщені, якщо пояснюючі змінні корелюють із помилками регресії в більш ранні моменти часу, але не корелюють у той самий момент часу;
- зміщені і неспроможні, якщо пояснюючі змінні і помилки регресії корелюють у тому числі і в однакові моменти часу.

В точності такі самі результати можуть бути отримані і у випадку множинної регресії, коли b є векторним параметром.

Отже, корельованість регресорів і помилок регресії виявляється значною мірою більш неприємною обставиною, чим, наприклад, гетероскедастичність або автокореляція. Неадекватними виявляються не тільки результати тестування гіпотез, але і самі оцінювані значення параметрів.

Розглянемо дві причини корельованості регресорів і помилок регресії, що зустрічаються найчастіше.

1. На випадковий член ε впливають ті самі фактори, що і на формування значень регресорів.

Нехай, наприклад, існує змінна u така, що в регресійних моделях

$$y_t = \alpha + \beta x_t + \gamma u_t + v_t, \quad (7.2)$$

$$x_t = \lambda + u_t + \xi_t, \quad (7.3)$$

коефіцієнти γ і δ значуще відрізняються від нуля. Вважатимемо для простоти, що є лише один регресор x , розглянутий як випадкова величина.

Припустимо, що ми не маємо спостережуваних значень u і розглядаємо модель

$$y_t = \alpha + \beta x_t + \varepsilon \quad (7.4)$$

Природно очікувати в цьому випадку корельованість регресорів x і помилок регресії ε .

Розглянемо приклад. В пункті A виробляють сировину двох видів I і II. Сировину перевозять до пункту B , де виробляють субпродукт, що продають фі-

рмі-виробникові за ціною X . Фірма виготовляє з субпродукту кінцевий товар, що перевозять до пункту C і реалізують за ціною Y . Ціни на сировину першого і другого типу змінюються і утворюють часові ряди Z_1 і Z_2 .

Побудуємо залежність ціни реалізації Y від ціни X , одержимо регресійну модель

$$y_t = \alpha + \beta x_t + \varepsilon_t.$$

Очевидно, що доставка сировини до пункту B і доставка кінцевого продукту до пункту C пов'язані з перевезеннями, а виходить, такі фактори, як витрати на паливо, зарплата водіїв, стан доріг і таке інше впливатимуть і на формування ціни X , і на кінцеву ціну Y при заданому X , тобто на величину помилок регресії моделі. Таким чином, регресори і помилки регресії виявляються корельованими, і оцінки, які отримані за методом найменших квадратів, будуть неспроможні.

Відзначимо другу типову причину корельованості пояснюючих змінних і випадкового члена.

2. Помилки при вимірі регресорів. Нехай при вимірі регресору x_j допускають випадкову помилку u_j , що задовольняє умові $M(u_{ij})=0$, тобто в обробку надходить не істинне спостережуване значення x_{ij} а перекручене

$$(x_{ij})^* = x_{ij} + u_{ij}. \quad (7.5)$$

При оцінюванні моделі фактично розглядають регресію

$$y = x\beta + v. \quad (7.6)$$

Користуючись методом найменших квадратів, саме з цього рівняння ми і дістанемо оцінку β , в той час як у дійсності, маємо:

$$y = x\beta + \varepsilon = (x^* - u)\beta + \varepsilon = x^*\beta + (\varepsilon - u\beta),$$

тобто $v = \varepsilon - u\beta$, і, отже, у фактично розглянутій моделі є кореляція між значеннями регресорів x^* , що є фіксованими, і випадковим членом v .

Відзначимо, що обидві зазначені причини корельованості регресорів і помилок регресії мають той самий математичний зміст: значення пояснювальних змінних формуються не присутнім в моделі регресором, а якимось іншим, і, отже, оцінюється «не той» параметр.

При розгляді конкретних регресійних моделей часових рядів з корельованістю регресорів і помилок доводиться зіштовхуватися досить часто. Проблему визначення спроможної оцінки параметру β дозволяють вирішити ряд розроблених для цього методів.

7.2 Методи оцінки параметрів з урахуванням лагових ефектів

Одним з методів визначення спроможної оцінки параметра β моделі (7.1) є метод інструментальних змінних. Він полягає в тому, що підбирають ряд нових змінних z_j ($j = 1, t$), які тісно корелюють з x_j і не корелюють з ε у рівнянні (7.1).

Якщо розглядають модель парної регресії, і єдину змінну x замінюють на одну інструментальну змінну z , формула спроможної оцінки параметра β має вигляд:

$$\beta = \frac{\text{cov}(z_t, y_t)}{\text{cov}(z_t, x_t)} = \beta + \frac{\frac{1}{n} \sum_{i=1}^n z_{it} \varepsilon_{it}}{\text{cov}(z_t, x_t)}. \quad (7.7)$$

Можна показати, що чим тісніше корелюють вихідні змінні x і інструментальні змінні z , тим ефективнішою буде оцінка β . При цьому, повинна виконуватися умова $\text{Cov}(z, \varepsilon)=0$. Отже, оптимальними інструментальними змінними є змінні виду

$$z_j^* = x_j - M_{\varepsilon}(x_j), \quad j=l, p,$$

де x, z - p -мірні випадкові величини.

Проте випадкові величини z^* є неспостережуваними, отже, найкращого набору інструментальних змінних реально не існує. Проблему вирішують, якщо існує довільний набір інструментальних змінних z_j , що мають реальний економічний зміст, причому їхнє число може перевершувати число вихідних регресорів x_j . Використовуючи МНК для визначення прогнозних значень x_j на підставі моделі

$$x_j = z \gamma + v,$$

потім використовують прогнозні значення x_j як інструментальні змінні. Змінні x_j не корелюють з помилками регресії, тому що лінійно виражаються через інструментальні змінні z_j .

Такий підхід до оцінки параметрів регресії називають двокроковим методом найменших квадратів. По суті метод найменших квадратів застосовують двічі: спочатку для одержання набору регресорів x , потім для одержання оцінок параметра β .

Процедура двокрокового методу найменших квадратів реалізована в більшості комп'ютерних регресійних пакетів.

На практиці, як правило, можливості вибору інструментальних змінних не широкі, а нерідко буває, що спостережувані інструментальні змінні відсутні.

В практичних задачах найчастіше зустрічається кореляція між $y_{t-\tau}$ і ε_t . В цьому випадку можна використати авторегресійну модель з розподіленими лагами порядків p і q виду

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \dots + \beta_p x_{t-p} + \gamma_1 y_{t-1} + \dots + \gamma_q y_{t-q} + \varepsilon_t. \quad (7.8)$$

Саме такого виду моделі мають найбільше практичне значення, і саме такого виду механізм виникнення кореляції між регресорами і помилками регресії найбільш часто зустрічається в економічних задачах. Серед них найчастіше виникають моделі порядку $(0,1)$, що мають вигляд:

$$y_t = \alpha + \beta x_t + \gamma y_{t-1} + \varepsilon_t \quad (7.9)$$

де помилки регресії ε_t підкоряються авторегресійному процесу першого порядку, тобто $\varepsilon_t = \rho \varepsilon_{t-1} + \xi_t$ і ξ_t є незалежними однаково розподіленими випадковими величинами з нульовими математичними сподіваннями і дисперсіями σ^2 .

Можна показати, що застосування до моделі (7.8) звичайного методу найменших квадратів дає оцінку параметру γ , що збігається за імовірністю до величини виду

$$\frac{\gamma + \rho}{1 + \rho}.$$

Неспроможність оцінки γ тим більша, чим сильніша автокореляція помилок ε . На практиці, проте, часто виконується умова $\rho \ll \gamma$. У цьому випадку границя оцінки, отриманої за методом найменших квадратів, буде близькою до дійсного значення параметру γ , хоча і не дорівнює йому.

Якщо умова $\rho \ll \gamma$ не виконується, звичайний метод найменших квадратів може давати істотне відхилення від дійсного результату навіть на вибірках великого обсягу. Для вирішення цієї проблеми діють у такий спосіб. Записують модель (7.9) для моменту часу $t-1$ і підставляють до виразу (7.9); далі записують рівняння (7.9) у момент часу $t-2$ і підставляють в отриманий вираз. Продовжуючи цей процес нескінченно, одержують вираз:

$$y_t = \frac{\alpha}{1 - \gamma} + \beta x_t + \beta \sum_{k=1}^{\infty} \gamma^k x_{t-k} + \sum_{k=0}^{\infty} \gamma^k \varepsilon_{t-k}. \quad (7.10)$$

Модель (7.10) називають моделлю з розподілом Койка лагових пояснюючих змінних. Її ще іноді називають моделлю з геометричним розподілом, маючи на увазі, що коефіцієнти при лагових змінних утворюють геометричну прогресію із знаменником γ ($\gamma < 1$). Перетворення моделі (7.9) до виду (7.10) називають зворотним перетворенням Койка.

Змінні x не корелюють з помилками ε , так що застосуванням зворотного перетворення Койка вирішують проблему корельованості регресорів з випадковими членами. Проте застосування звичайного методу найменших квадратів до моделі (7.10) виявляється на практиці неможливим через нескінченно велику кількість регресорів. Через те, що коефіцієнти, які входять до моделі ряду, зменшуються у геометричній прогресії, і сам ряд швидко збігається, можна було б обмежитися порівняно невеликим числом лагів. Проте у цьому випадку з'являються, принаймні, дві важко розв'язуваних проблеми. По-перше, виникає сильна мультиколінеарність, тому що лагові змінні сильно корельовані. По-друге, рівняння виявляється неідентифікованим.

Рівняння (7.10) можна оцінити за допомогою процедури, яку називають нелінійним методом найменших квадратів. Сутність нелінійного методу найменших квадратів збігається до наступного.

З досить дрібним кроком (наприклад, 0,01) перебирають всі значення γ з можливої області значень цього параметра (якщо ніякої апріорної інформації

немає, то ця область - інтервал $(0, 1)$. Одержують послідовність значень $\gamma^{(a)}$. Потім для кожного значення $\gamma^{(a)}$ обчислюють значення

$$x_t^{(a)} = x_t + \sum_{k=1}^k (\gamma^{(a)})^k x_{t-k}.$$

Границю підсумовування k обирають так, що наступні члени ряду вносять у його суму незначний внесок (очевидно, чим менше обране значення $\gamma^{(a)}$, тим менше членів ряду доводиться враховувати). Для кожного $x^{(a)}$ за методом найменших квадратів одержують параметри рівняння $y_t = \alpha + \beta x_t^{(a)} + u_t$, а потім обирають те рівняння, що забезпечує найбільший коефіцієнт детермінації R^2 . Відповідне значення $\gamma^{(a)}$ приймають за оцінку параметру γ і обчислюють оцінки α і β , потім знаходять оцінки вихідних параметрів.

Процедура нелінійного методу найменших квадратів реалізована в більшості комп'ютерних пакетів.

Зауважимо на те, що хоча за допомогою зворотного перетворення Койка усувається корельованість регресорів з помилками, але автокореляція помилок здобуває складну структуру, і усунення її може виявитися практично неможливим.

В економічних задачах так само часто зустрічається ситуація, коли під впливом пояснюючої змінної x формується не сама величина y , а її ідеальне, «бажане» (desired) значення y^{des} . Реальне ж спостережуване значення y , являє собою зважену суму бажаного значення в момент часу t і спостережуваного в попередній момент $t-1$.

Нехай, наприклад, якась фірма виплачує в момент часу t дивіденди y_t . Природно вважати, що сума дивідендів являє собою певну частку прибутку фірми x_t :

$$y_t = \lambda x_t. \quad (7.11)$$

На практиці рівняння (7.11) піддають частковому коректуванню. Якщо прибуток виявиться малим, тоді на частку дивідендів піде більша частина, ніж y , оскільки відомо, що зменшення дивідендів завдає серйозного удару престижу фірми. З цієї самої причини у випадку великого прибутку частка дивідендів виявиться меншою - фірма виявить обережність: можливо, у майбутньому періоді прибуток зменшиться, і тоді доведеться урізати дивіденди (іншим фактором, що стримує ріст дивідендів може послужити бажання фірми інвестувати частину прибутку в розширення виробництва). В результаті реальну зміну дивідендних виплат Δy_t , формують в такий спосіб:

$$\Delta y_t = \lambda(y_t^{des} - y_{t-1}) + v_t, \quad 0 < \lambda < 1. \quad (7.12)$$

Рівняння (7.12) називають рівнянням часткового коригування. Рівняння (7.12) разом з рівнянням (7.11) дають наступну модель:

$$\Delta y_t = \lambda(y_t^{des} (1 - \lambda) y_{t-1} + v_t, \quad (7.13)$$

$$y_t^{des} = \alpha + \beta x_t.$$

Модель (7.13) називають моделлю часткового коригування. (Для моделі виплати дивідендів можна вважати, що $\alpha = 0$; у загальному ж випадку вільний член присутній у рівнянні формування бажаного значення). Тут важливо відзначити, що величина y_t^{des} є неспостережуваною.

Очевидно, регресійне рівняння моделі (7.13) можна записати у вигляді:

$$y_t = \alpha\lambda + \beta\lambda x_t + (1 - \lambda)y_{t-1} + v_t, \quad (7.14)$$

воно являє собою авторегресійну модель з розподіленими лагами порядку (0,1) і може бути оцінене за нелінійним методом найменших квадратів після зворотного перетворення Койка. Відзначимо, що спроможні оцінки параметрів рівняння (7.14) можна одержати й за звичайним методом найменших квадратів, оскільки в рівнянні пояснююча змінна y_{t-1} не корелює із значенням випадкового члену \mathcal{E} у момент часу t .

Ще одним важливим прикладом регресійної моделі з розподіленими лагами є модель адаптивних сподівань.

Нехай $y_t = \log \frac{m_t}{P_t}$, де m - номінальна кількість грошей в обігу, P - рівень

цін. Величину $\frac{m}{P}$ називають реальними грошовими залишками. Нехай y_t^d - попит на реальні грошові залишки.

Ф. Кейган, вивчаючи динаміку цієї величини в періоди гіперінфляції, висунув припущення, що її значення в момент t визначається очікуваним рівнем інфляції в момент $t+1$. Модель, запропонована Кейганом, має вигляд:

$$y_t = \alpha + \beta x_t^w + \varepsilon_t, \quad (7.15)$$

де x^w - очікуваний рівень інфляції.

Очевидно, це неспостережувана величина. Кейган доповнив рівняння (7.15) рівнянням

$$x_{t+1}^w = \lambda(x_t - x_t^w), \quad (7.16)$$

де $x_{t+1}^w = x_t - x_t^w$.

Перепишемо рівняння (7.15) і (7.16) у вигляді:

$$x_{t+1}^w = \lambda x_t + (1 - \lambda)x_t^w \quad (7.17)$$

$$y_t = \alpha + \beta x_{t+1}^w + \xi_t. \quad (7.18)$$

Тепер стає зрозумілим зміст рівняння (7.16), який полягає в тому, що очікуваний рівень інфляції в момент $t+1$ являє собою зважену суму очікуваного й реального рівня в момент t .

Модель (7.18) називають моделлю адаптивних очікувань. (Модель гіперінфляції Ф. Кейгана, очевидно, являє собою вперше розглянутий приклад такої

моделі.) З рівнянь (7.15) і (7.16) може бути виключена неспостережувана величина x^w . Запишемо рівняння (7.16) у момент часу $t-1$:

$$y_{t-1} = \alpha + \beta x_t^w + \xi_{t-1}.$$

У той самий час маємо:

$$\begin{aligned} y_t &= \alpha + \beta(\lambda x_t + (1-\lambda)x_t^w) + \xi_t = \\ &= \alpha + \beta\lambda x_t + (1-\lambda)\beta x_t^w + \xi_t = \\ &= \alpha + \beta\lambda x_t + (1-\lambda)(y_{t-1} - \alpha - \xi_{t-1}) + \xi_t \end{aligned}$$

Остаточно одержимо:

$$y_t = \alpha\lambda + \beta\lambda x_t + (1-\lambda)y_{t-1} + \varepsilon_t, \quad (7.19)$$

де

$$\varepsilon_t = \xi_t - (1-\lambda)\xi_{t-1}. \quad (7.20)$$

Модель (7.19) є авторегресійною моделлю з розподіленими лагами порядку (0,1). Явний вигляд (7.20) випадкового члену показує, що має місце кореляція між лаговою змінною y_{t-1} і помилкою регресії ε_t , тобто оцінки методу найменших квадратів не будуть спроможними.

Модель (7.19) можна оцінити, застосувавши зворотне перетворення Койка і потім нелінійний метод найменших квадратів.

Найбільш відомим прикладом моделі адаптивних сподівань є модель споживання М. Фрідмена. Розглянемо її докладніше.

Вивчаючи залежність між споживанням і доходом індивідумів, М. Фрідмен припустив, що пропорційна залежність повинна будуватися не між фактичними величинами, а між їхніми постійними складовими. Постійний дохід - це сума, на яку людина може розраховувати у відносно довгостроковий період (заробітна плата, стабільні гонорари, відсотки з внесків і т.д.). Фактичний дохід у розглянутий момент часу може значно відрізнятись від постійного. Аналогічно, постійне споживання - це, по суті, звичний рівень споживання. Його фактичне значення може сильно відрізнятись від постійного у випадку великої покупки або непередбачених витрат.

М. Фрідмен виходив з припущення, що постійне споживання індивіда y^c пропорційне його постійному доходу x^c , тобто

$$x_t^c = \beta x_t^c. \quad (7.21)$$

В той же час фактичне споживання Y і фактичний дохід X являють собою суму постійних і часових величин:

$$Y = Y^c + Y^T, \quad X = X^c + X^T, \quad (7.22)$$

причому часові величини X^T , Y^T є випадковими. Гіпотеза Фрідмена полягала в тому, що ці величини не корелюють в різні моменти часу і між собою, їхні математичні сподівання дорівнюють нулю, а дисперсії постійні у часі.

Підставивши вираження (7.21) в (7.22), одержимо регресійну модель

$$y_t = \beta x_t^c + y_t^T. \quad (7.23)$$

Проблема полягає в тому, що постійні дохід і споживання є суб'єктивними, а отже, неспостережуваними величинами. М. Фрідмен застосував до вивчення залежності (7.23) модель адаптивних сподівань, припустивши, що зміна постійного доходу пропорційна різниці між його реальним значенням і попереднім постійним значенням, тобто

$$\Delta x_t^C = \lambda(x_t - x_{t-1}^C). \quad (7.24)$$

Це означає, що при збільшенні реального доходу індивіди корегують своє уявлення про постійний дохід, але не на повне значення приросту, а на певну його частину, розуміючи, що приріст може виявитися обумовленим тимчасовою, тобто випадковою, складовою. Рівняння (7.24) можна записати в стандартній формі моделі адаптивних сподівань:

$$x_t^C = \lambda x_t + (1 - \lambda)x_{t-1}^C, \quad (7.25)$$

або у вигляді наступного рівняння:

$$y_t = \beta \lambda x_t + (1 - \lambda)x_{t-1} + \varepsilon_t, \quad (7.26)$$

де $\varepsilon_t = y_t^T - (1 - \lambda)y_{t-1}^T$, звідки видно, що в моделі (7.25) має місце корельованість регресора Y_{t-1} з випадковим членом.

Модель (7.25) можна оцінити за допомогою нелінійного методу найменших квадратів. Також до моделі (7.23) можна застосувати метод інструментальних змінних. Вперше це було зроблено Н. Левітаном, що скористався як інструментальними змінними для X^C фактичними доходом і споживанням на іншому часовому відрізку.

Запитання для самоперевірки

1. Наведіть приклади економічних завдань, для яких економетричне моделювання потребує використання моделей з розподіленим лагом та моделей авторегресії.
2. Що в економіці називають лагом? Що є причиною лага?
3. Які регресійні моделі називають авторегресійними?
4. Які проблеми виникають під час оцінювання параметрів авторегресійних моделей?
5. Які гіпотези висувають щодо залишків лагових моделей?
6. Охарактеризуйте інтерпретацію параметрів моделі з розподіленим лагом. Перелічіть абсолютні та відносні чинники сили зв'язку моделі з розподіленим лагом.
7. Охарактеризуйте інтерпретацію параметрів моделі авторегресії.

СПИСОК ДЖЕРЕЛ

1. Наконечный С. И. Эконометрия : учебник / С. И. Наконечный, Т. П. Терещенко. - К.: КНЕУ, 2001.
2. Кремер Н. Ш. Эконометрика : Учебник для вузов / [Н. Ш. Кремер, Б. А. Путко; под ред. проф. Н. Ш. Кремера]. - М.: ЮНИТИ-ДАНА, 2002. - 311 с.
3. Практикум по эконометрике : Учебн. пособие / [И. И. Елисеева, С. В. Курешева, Н. М. Гордиенко и др.; Под ред. И. И. Елисеевой]. - М.: Финансы и статистика, 2002. - 192 с.
4. Замков О. О. Математические методы в экономике / О. О. Замков. - М.: Финансы и статистика, 2001.
5. Магнус Я. Р. Эконометрика. Начальный курс / Я. Р. Магнус, П. К. Катышев, А. А. Пересецкий. - М.: Дело, 2001. - 400 с.
6. Кулинич Е. И. Эконометрия : учеб. пособ. / Е. И. Кулинич - М.: Финансы и статистика, 2000. - 304с.
7. Афанасьев В. Н. Анализ временных рядов и прогнозирование : Учебник / В. Н. Афанасьев, М. М. Юзбашев. - М.: Финансы и статистика, 2001. - 228 с.
8. Экономико-математические методы и модели : Учебн. пособие / Н. И. Хлод, А. В. Кузнецов, Я. Н. Жихар и др.; Под общ. ред. А. В. Кузнецова. 2-е изд. - Мн.: БГЭУ, 2000. - 412 с.
9. Экономико-математические методы и прикладные модели : Учебн. пособие для вузов / В. В. Федосеев, А. Н. Гармаш, Д. М. Дайитбеков и др.; Под ред. В. В. Федосеева. - М.: ЮНИТИ, 2001. - 391 с.
10. Доугерти К. Введение в эконометрику. - М.: Финансы и статистика, 1999. - 402 с.
11. Лещинский О. Л. Эконометрия: учеб. пособ. / О. Л. Лещинский, В. В. Рязанцева, О. О. Юнькова. - К.: МАУП, 2003.

Навчальне видання

ВОРОНКОВ Олексій Олександрович

КОНСПЕКТ ЛЕКЦІЙ

з курсу

ЕКОНОМЕТРИКА

(для студентів заочної форми навчання освітньо-кваліфікаційного рівня «бакалавр» напрямку підготовки 6.030504 - Економіка підприємства)

Відповідальний за випуск: *А. Є. Ачкасов*

За авторською редакцією

Комп'ютерне верстання: *І. В. Волосожарова*

План 2015, поз. 182 Л

Підп. до друку 31.03.2015 р.
Друк на різнографі
Тираж 50 пр.

Формат 60х84/16
Ум. друк. арк. 5,3
Зам. №

Видавець і виготовлювач:
Харківський національний університет
міського господарства імені О. М. Бекетова,
вул. Революції, 12, Харків, 61002
Електронна адреса: rectorat@kname.edu.ua
Свідоцтво суб'єкта видавничої справи:
ДК № 4705 від 28.03.2014 р.